

Communication, Management and Information Technology

Editor: Marcelo Sampaio de Alencar

COMMUNICATION, MANAGEMENT AND INFORMATION TECHNOLOGY



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON COMMUNICATION,
MANAGEMENT AND INFORMATION TECHNOLOGY (ICCMIT 2016), COSENZA, ITALY, 26–29
APRIL 2016

Communication, Management and Information Technology

Editor

Marcelo Sampaio de Alencar

Institute for Advanced Studies in Communications (Iecom), Campina Grande-PB, Brazil



CRC Press

Taylor & Francis Group

Boca Raton London New York Leiden

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

A BALKEMA BOOK

CRC Press/Balkema is an imprint of the Taylor & Francis Group, an informa business

© 2017 Taylor & Francis Group, London, UK

Typeset by V Publishing Solutions Pvt Ltd., Chennai, India

Printed and bound in Great Britain by CPI Group (UK) Ltd, Croydon, CR0 4YY.

All rights reserved. No part of this publication or the information contained herein may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, by photocopying, recording or otherwise, without written prior permission from the publisher.

Although all care is taken to ensure integrity and the quality of this publication and the information herein, no responsibility is assumed by the publishers nor the author for any damage to the property or persons as a result of operation or use of this publication and/or the information contained herein.

Published by: CRC Press/Balkema

P.O. Box 11320, 2301 EH Leiden, The Netherlands

e-mail: Pub.NL@taylorandfrancis.com

www.crcpress.com – www.taylorandfrancis.com

ISBN: 978-1-138-02972-9 (Hbk)

ISBN: 978-1-315-37508-3 (eBook PDF)

Table of contents

Preface	xi
Organizers	xiii
Critical discourse analysis on inclusive education: School leadership and networks <i>R. Soler Costa & J.R. Soler Santaliestra</i>	1
An empirical study of the MIS impacts and organizational efforts on improving business performance <i>A.A. Monem</i>	11
Social benefits of innovative public administration services <i>B. Skoczyńska-Prokopowicz</i>	17
The use of Information and Computer Technology in the decision-making process of foreign students concerning their future professional career <i>B. Sobieczewska</i>	23
Modelling of the process teaching-training in E-learning <i>A.T.A. Yalid, M. Bassiri, M. Moussted & M. Talbi</i>	29
The instrumentalisation of the MOOCS vector of educational innovation and consecration of the academic training quality <i>A.T.A. Yalid, M. Bassiri, M. Moussted & M. Talbi</i>	33
Interdisciplinary as a vector of consecration and development metacognitive transversal on line skills <i>B. Mustapha, Y. Amal, S. Belaouad & R. Mohamed</i>	37
Interactive formation from afar and cognitive intelligence increased at the adult learners: Case of the master technological engineering for the education and the formation <i>B. Mustapha, Y. Amal, S. Belaouad & R. Mohamed</i>	45
The engineering andragogical in the device “blended learning”: Case of the academic formation courses professionalizing “development of a conceptual and methodological” setting <i>B. Mustapha, Y. Amal, S. Belaouadb & R. Mohamed</i>	53
A general scheme for MRI images’ segmentation based on slice-by-slice learning <i>I. Kone, L. Boulmane, M. Radouani & B. Elfahime</i>	61
FreebaseViz: Interactive exploration of freebase schema using query-driven visualisation <i>M. Elbattah, M. Roushdy, M. Aref & A.-B.M. Salem</i>	65
Towards a future mobile multihomed environment <i>A. Benaouda Chaht, C. Zouaoui & A. Bounoua</i>	73
Information security challenge: The responsibility of management, Information System case study for the management of research <i>Y. El Hissi & A. Haqiq</i>	79
Effects and impact of playing computer games <i>W. Chmielarz & O. Szumski</i>	87

The utilization of the HR analytics by the high and mid-level managers: Case from Eastern Poland <i>M. Wawer & P. Murymas</i>	97
The relation between friendship, commuting time, and student performance: A Social Network Analysis <i>L.J. Khalil & M. Khair</i>	107
Optimized clustering protocol for Wireless Sensor Networks using compressive sensing <i>D.M. Omar & A.M. Khedr</i>	115
Compositional writing to cope with electronic media <i>F.M. Sadek</i>	125
MVDR beamformer model for array response vector mismatch reduction <i>S.N. Shahab, A.R. Zainun, E.I. Essa, N.H. Noordin, I.I. Mohamed & A. Omar Khaldoun</i>	131
Enhancing quality of service by balancing the traffic load in Mobile Ad hoc Networks (MANETs) <i>G. Amina & A. Boukelif</i>	139
Artificial intelligence in e-learning <i>H. Alaoui Harouni, E. Hachem & C. Ziti</i>	145
Highlighting the evaluation gaits in the adaptive learning system ALS_CORR[LP] <i>N. El Ghouch, E.M. En-Naimi, Y.Z. Seghroucheni, B.E. El Mohajir & M. Al Achhab</i>	151
Towards a blended learning using mobile devices, podcasts and QR codes in Algeria <i>S. Ghizlene, K. Belkacem & D. Mohamed</i>	159
Evolutionary algorithm to solver impairment aware wavelength assignment problem <i>A.M.L. Miranda, C.A.J. Rocha, J.C.W.A. Costa & C.W.A. Costa</i>	167
Using the 3M method for the optimization of the managerial act <i>B.-A. Furduescu</i>	171
Business Process modeling: Case of undergraduate program <i>K.V. Zhukova & A. Yu. Pleshkova</i>	179
A full duplex media access protocol for hybrid visible light communication networks <i>X. Wang, L. Zhang & W. Dou</i>	187
Combining feature extraction methods to classify three motor imagery tasks <i>M.H. Zaky, A.A. Nasser & M.E. Khedr</i>	195
Telemedicine home program in patients with cystic fibrosis: Results after 10 Years <i>F. Murgia, I. Tagliente, V. Mercuri, S. Bella, F. Bella, I. Zoppis, G. Mauri & F. Sicurello</i>	203
DIABESITY: Design of mHealth integrated solutions for empowering diabetic and obese citizens in self-monitoring and self-management using mobile devices, apps, social media and web-based technologies <i>I. Zoppis, G. Mauri, F. Sicurello, E. Santoro & G. Castelnuevo</i>	207
Mining complex networks: A new challenge for supporting diagnostic decisions <i>I. Zoppis, G. Mauri, F. Sicurello, E. Santoro & G. Castelnuevo</i>	215
VoIP providers trunk round trip delay remote measurement methodology <i>M. Mikulec, J. Rozhon & M. Voznak</i>	221
Automatic identification and data capture techniques by radio frequency identification RFID tags applied to reader authentication <i>H. Saadi, R. Touhami & M.C.E. Yagoub</i>	227
The impact of using educational gamification in mobile computing course: A case study <i>R. Al-Azawi, M. Al-Obaidy, A. Ayesh & D. Rosenburg</i>	235
Efficient mining of high average-utility itemsets <i>J.C.-W. Lin, T. Li, P. Fournier-Viger, T.-P. Hong & M. Voznak</i>	241

An extension for the mobile payment collaborative model proposed for developing countries—Egypt case study <i>M. Goher & M.A. Rizka</i>	249
Measuring adolescents awareness security of internet a critical analysis of the internet and adolescents self-injury <i>F. Ben Salamh</i>	257
Optimal throughput of time power switching relaying protocol with imperfect channel state information <i>H.-S. Nguyen, D.-T. Do, A.-H. Bui Thi & M. Voznak</i>	261
Content based video retrieval—towards full multimedia join <i>R.A. Abinader & P.G. Gedeon</i>	269
Modeling the seller’s strategy when sales depend on buzz <i>O. Lefebvre</i>	279
Verification of merge sorting technique using formal methods <i>N.M. Zaitoun & M.J. Agel</i>	287
Barriers surrounding e-government implementation: A case study of Government to Business (G2B) system <i>H. Hashim, A. Lin & J. Foster</i>	295
The effectiveness of an educational website for promoting design skills and use of educational blogs by teachers of secondary education in Saudi Arabia <i>M.A.S. Al Mozaini</i>	299
Using machine learning technique towards personalized mobile flyer <i>M.A. Razeq & H.G. Bardessi</i>	305
Barriers of knowledge acquisition in telecommunications companies in Saudi Arabia: An exploratory study on Etihad Etisalat Mobily <i>M.M. Al-Harithy</i>	313
The effect of using computerized educational software based on interactive video in developing some of using computer skills for the preparatory year students at Albaha University, Saudi Arabia <i>R.S. Abdullah AlSelouly</i>	321
The reality of tacit knowledge sources in educational training center in department of education of Bisha province: Saudi Arabia case study <i>A.M.A. Alyateem</i>	327
Digital institutional repositories in Saudi universities between facts and future: A comparative study between Australian and Saudi digital institutional repositories <i>K. Bawazeer</i>	335
The role of digital repositories in supporting the main functions of universities: A survey of Arab universities <i>N.J. AlKhudairi</i>	339
Importance of integrating knowledge management methods and tools to enhance risk management processes. Exploratory study in Saudi Arabia business environment <i>S.S. Humaidan</i>	345
Supporting knowledge society using digital repositories <i>E.H. ALZahrani</i>	353
Impact of information resources on decision-making process in different enterprises <i>H.M. Albogami</i>	359
The reality of using various information resources to support the managerial decision-making process: Survey study on Saudi Arabian airlines enterprise in Jeddah City <i>H.M. Albogami</i>	365

Brain storming algorithm for coverage and connectivity problem in wireless sensor network <i>R.A. Ramadan & A. Y. Khedr</i>	371
Evaluating IPTV network performance using OPNET <i>E.S. Sabry, R.A. Ramadan, M.H. Abd El-Azeem & H. ElGouz</i>	377
Time Based Weighted Shortest Path Movement Model (TBW-SPMM) <i>A.B. Altamimi</i>	385
Video steganography using P-frame motion vector <i>A.E. Ibrahim, M.A. Elshahed & T.I. Elarif</i>	389
A study to evaluate the role of social networks in supporting the educational process: Survey study <i>A. Almajhaddi & S. Almutairi</i>	395
Promote the process of higher education based on social media as a creative ICT tools <i>A. Almajhaddi & S. Almutairi</i>	405
ICT and digital sources challenges in the university library: The case of King Khalid University library <i>S.Q. Al-Khalidi Al-Maliki</i>	417
An exploratory evaluation of the awareness of e-government services among citizens in Saudi Arabia <i>S.Q. Al-Khalidi Al-Maliki</i>	425
The role of social media in creating new knowledge for graduate students at King Abdulaziz University in Jeddah, Saudi Arabia <i>S.A. Al-Afghani & H.M. Albogami</i>	433
Fuzzy decision trees for text document clustering <i>W.B. Abdessalem, K. Dridi & E. Alkhammash</i>	439
Generation of use case UML diagram from user requirement specifications <i>W.B. Abdessalem & E. Alkhammash</i>	447
Modeling guidelines of FreeRTOS in Event-B <i>E. Alkhammash, M. Butler & C. Cristea</i>	453
Comparative study of various open source digital content management software <i>F.A. Al Selami</i>	463
Institutional digital repositories specializing in cognitive science: An exploratory study <i>F.A. Al Selami</i>	471
Enterprise architecture moving from professional certificates into academic credentials <i>F. Fouad</i>	477
Digital warehouse for research in economics: Analytical study between reality and expectations <i>E.H. ALZahrani</i>	483
Social network applications: Critical review study <i>M.M. Abu Sharha</i>	491
The effective use of social networks to support knowledge acquisition process at KAU, Jeddah, Saudi Arabia <i>M.I. Yousef & E.H. ALZahrani</i>	497
The role of open access in supporting the digital repositories activities <i>M.I. Yousef</i>	503
The digital repository Arxiv: A comparative study with similar repositories <i>M.I. Yousef</i>	511
Pros and cons of social networks and their impact on human behaviour <i>A.A. Alyoubi & I.M. Alharbi</i>	517

Scientific and traditional communication obstacles of free access to information in institutional digital repositories <i>A.S.H. Al-Msloum, A.A. Al-Johani & O.A.A. Alsulami</i>	527
Assessment on course learning outcome aligned to students' achievement <i>A.B. AbdulAziz</i>	533
Management aspects of big data in various enterprises <i>B.A. Alyoubi & I.M.M. El Emary</i>	537
Towards an approach based on hadoop to improve and organize online search results in big data environment <i>K. Aoulad Abdelouarit, B. Sbihi & N. Aknin</i>	543
The efficacy of using electronic educational bag in developing functional expression skills among secondary school students at Bisha Governorate, Saudi Arabia <i>M.S.A. Al Mosaar</i>	551
A proposed real-time EMG intelligent robot control algorithm based on swarm intelligence and neural networks <i>B.M. ElBagoury, J. Al-Amri & M. Roushdy</i>	557
Solving the problems of linguistically diverse the 1st year university student's using digital learning <i>D. Ratniece & S. Cakula</i>	565
Multi-agents based framework for selecting cloud service provider <i>M. Abo-Rizka & R. El-Awadi</i>	573
Human work perspectives in cyber-physical systems impacting industry & business (<i>with accent on Czech Republic</i>) <i>E. Kasparova</i>	583
Statistical analysis of mobility patterns and passive bandwidth reservation in vehicular networks based on a dynamic programming approach <i>M. Tropea & F. Strangis</i>	589
WARM in the city: WAste Route Management in the smart city (WARM City) <i>M. Tropea, A.F. Santamaria & S. Marano</i>	597
A new application for analyzing driving behavior and environment characterization in transportation systems based on a fuzzy logic approach <i>P. Fazio, A.F. Santamaria, M. Tropea, A. Serianni & F. Cirillo</i>	599
Trust-based intrusion detection in mobile ad-hoc networks using a dynamic approach for energy-efficient monitoring <i>A. Lupia</i>	609
Anemia types prediction based on data mining classification algorithms <i>M. Abdullah & S. Al-Asmari</i>	615
Energy Efficient Optimized Routing Algorithm (EEORA) <i>N. Alharbe & M. Abdullah</i>	623
MERS-CoV Disease Estimation (MDE) A study to estimate a MERS-CoV by classification algorithms <i>M. Abdullah, M.S. Altheyab, A.M.A. Lattas & W.F. Algashmari</i>	633
E-learning standards <i>M. Abdullah & N. Abdel Aziz Ali</i>	639
Cross-layer quality of service protocols for wireless multimedia sensor networks <i>A. AlAmri & M. Abdullah</i>	649
OXLP: An optimized cross-layers protocol for wireless sensor networks <i>A.S. Althobaiti & M. Abdullah</i>	659

Classification for data stream clustering protocols in wireless sensor networks <i>Y. Alghamdi & M. Abdullah</i>	671
An architecture for selling internet data using mobile hotspot <i>M. Taileb, B. Alshuaibi, W. Bagais, A. Basudan, N. Bahurmoz & M. Alsadi</i>	681
Big data mining: A classification perspective <i>N.M. Alotaibi & M.A. Abdullah</i>	687
ICT drivers of intelligent enterprises <i>M. Łobaziewicz</i>	697
The presence of sustainable entrepreneurship in Polish companies based on the selected examples <i>P. Bajdor</i>	703
Raising demand for implementation of systemic logistics management activities in agribusiness <i>A. Brzozowska & K. Szymczyk</i>	711
Implementation of the principles of the process orientation in the aspect of logistic management of supply chain <i>D. Bubel</i>	717
Analytical grounds and effective operation area of green logistics management in the transport area <i>M. Kadhubek & K. Grondys</i>	725
Information and communication support for the agricultural sector of Ukraine <i>A. Kalinichenko & O. Chekhlatyi</i>	731
Vendor Managed Inventory—implementation of VMI concept from the dynamic management perspective <i>H. Kościelniak & M. Starostka-Patyk</i>	737
Logistics chain management elements at global market of Liquefied Natural Gas (LNG) <i>M. Zawada & M. Starostka-Patyk</i>	745
An IoT course for a computer science graduate program <i>X. Liu & O. Baiocchi</i>	751
An innovative knowledge discovery mechanism for unique pattern <i>K. ElBahnasy</i>	757
The degree of knowledge management practice in the department of legal affairs: Saudi Arabian Airlines case study <i>A.M.A. Al-Yateem & N.B. bin Hamid</i>	765
A novel adaptive e-learning model matching educator-student learning styles based on machine learning <i>M. Abdullah, A. Y. Bayahya, E.S. Ba Shammakh, K.A. Altuwairqi & A.A. Alsaadi</i>	773
Feasibility for a seamless integration of admission, registration and academic advising in KAU-SIS <i>F. Fouad</i>	783
Author index	789

Preface

The International Conference on Communications, Management, and Information Technology (ICCMIT) is a biennial meeting seeks to provide a discussion forum for scientists, engineers, educators and students about the latest discoveries and realizations in the foundations, theory, models and applications of systems inspired on nature, using computational intelligence methodologies, as well as an emerging areas related to the three tracks of the conference. We strongly emphasize the wide range of topics comprised under the umbrella of (ICCMIT'2016), covering all the fields in communication Engineering, Management, and Information Technology. This conference (ICCMIT'2016) aimed to have significant contributions to various topics in Communications Engineering, Management, and Information Technology. The conference also included tutorials, workshops, and technology panels given by world-class speakers.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Organizers

 <p><i>Universal Society for Applied Research Czech Republic</i></p>	 <p><i>The University of Calabria (UNICAL) Cosenza -Italy</i></p>
<p>ITT Technical Institute ITT</p> <p><i>The ITT Technical Institutes - USA</i></p>	 <p><i>Institute of Advanced Studies in Communications-Brazil</i></p>

CONFERENCE CHAIR



Prof. Marcelo Sampaio de Alencar
*President of the Institute for Advanced Studies
in Communications (Iecom), Campina Grande-PB, Brazil*

CONFERENCE CO-CHAIR



Prof. Jacoub Saleh, Ph.D, MSEE, BSEE
*Chair, School of Electronics Technology,
ITT Technical Institute, USA*

CONFERENCE LOCAL CHAIR



Prof. Peppino Fazio
DIMES Department, University of Calabria, Italy



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Critical discourse analysis on inclusive education: School leadership and networks

Rebeca Soler Costa & Juan Ramón Soler Santaliestra

Faculty of Education, University of Zaragoza, Spain

ABSTRACT: School inclusion has caught up the stakeholders' attention when approaching compulsory school periods. Mainly this is due to the fact there exist an increasing diversity in students, so schools must offer the necessary measures to deal with this diversity. Assuming different measures have been introduced in the Education Acts in Spain, along this paper we are going to analyse those measures and value the degree of efficiency they have. It is obvious diversity refers not just to students with different ethnic backgrounds but also with immigrants who come to Spain and do not speak our native language and other students who may have special education needs. From a diachronic perspective, until 2006, the curricular prescriptions were shortly introduced, such as in the General Act of Education (1970), Organic Act Regulating the Right to Education (1985), Act for the General Organisation of the Education System (1990) and Organic Act of Participation, Evaluation and Government Schools (1995). Rather, if we focus our attention on the current Education Acts (Organic Act 2/2006 of 3d May of Education and Organic Act 8/2013 of 9th December for the Improvement of Quality in Education), we can observe school inclusion has received a great development. However, schools do not have the same opinion because they consider the amount of students' diversity is huge and the resources available to be applied are not enough. For this reason, it is necessary and interesting to analyse how school inclusion can be best developed in Primary and Secondary Compulsory Education. Consequently, we will go further with the analysis of the stakeholders' prescriptions and even we will restrict our attention to daily life school circumstances with the main aim of contributing to provide those students the educative attention they need. Thus, we will reach a real school inclusion. Obviously, as conclusions suggest, it is necessary to offer teachers a constant and updated training.

Keywords: School Organization, Didactics, assessment, methodology, school inclusion, curricular prescriptions' groupings, didactic strategies

1 INTRODUCTION

Roughly speaking, the implementations imposed by stakeholders through the different Education Acts have strong effects on the development of the teaching-learning processes. More precisely, in Spain, the still current Organic Act of Education (2006) offers significant conceptual changes in our education system. However, they are no longer a mere tautology. Going further with the conceptual categorization of Special Education in Primary and Secondary Compulsory school periods, it seems to be clear that students with specific educational needs require special human and material resources schools must offer them.

Language immersion programs and education enhancement programs help Students with Specific Need of Education Support who incorporate late, who have serious linguistic and cognitive shortcomings or who lack sufficient knowledge of

Spanish. They have been recently developed and are being introduced in Primary and Secondary classrooms. Similarly, Therapeutic Pedagogy specialists and counsellors provide individual attention to students who require a different teaching.

The concepts of inclusion and integration have been taken into consideration by any education discourse for a long time. However, most of the time, students with specific education needs are segregated from the rest of the class group. The principles of inclusion, integration, normalization and equity are present in any pedagogy speeches and legislative education texts. Authors like Stainback, Stainback and Jackson already showed in 1999 the conceptual change that was introduced in Spain in 2006. This change replaced the old term of *integration* with the more current term of *inclusion*. The reasons were varied. On the one hand, *inclusion* denotes more precisely the need to include all students in different social and educational tasks developed in schools.

On the other hand, *integration*, created by the Act for the General Organisation of the Education System (1990) was implemented and indeed involved the reinstatement of students at school, accepting the previous exclusion that some had suffered. Subsequently, the expression of *inclusive schools* was introduced, a concept that refers to the consolidation of an education system susceptible to include each and every student by offering individualised educational support. Integration implies the need to adapt to students who have been previously excluded returning them to the average groups. Inclusive educational responsibility lies within the professional development of teachers in charge of students with specific education needs in an attempt to satisfy their requirements.

2 INCLUSIVE EDUCATION IN A NATIONAL AND REGIONAL PERSPECTIVE

Stainback, Stainback and Jackson (1999) indicate that the change created is not only verbal but also conceptual. Under the expression of *inclusive schools* there are other priority parameters such as facing the needs of all students, not just those diagnosed as children with specific educational needs. The trends in the field of Education aim at creating a sense of community, an awareness and mutual support to promote the success of all students, by providing equal opportunities for all. There are other concerns in the development of school inclusion. Gartner and Lipsky (1987) and Stainback and Stainback (1990, 1992) show special interest in determining the type of work required to develop an appropriate inclusion.

They are the defenders of the ethical paradigm and are interested in the issues required to offer an inclusive education. Their interest lies not in ensuring the success of students with special education needs, but it is rather to ensure that all students are a part of a group class, regardless of their abilities, interests, skills, attitudes, family origins, etc. The goal of the ethical paradigm is to consider that inclusion is the fairest way where all students should be treated with dignity, without having to adapt to specific patterns or be subjected to the standards of the institution. Rather, the ethical paradigm must contribute to the students' diversity. Inclusion is a basic right all students must be offered, it is not a privilege. In the context of the ethical paradigm, Stainback and Stainback (1990) identify three reasons that support the creation and development of the inclusive school. Firstly, to give each student the chance to learn to live and work with their peers as something natural that happens in real life. This means that they are integrated into

an educational environment and community. Secondly, they aim to eliminate the inherent effects of segregation when children are placed in separate rooms, for example schools and/or special education rooms. Finally, to do what is fair, ethical and equitable.

These principles that encourage the development of such schools allow, according to the authors studied, a set of advantages over traditional approaches that try to help students with disabilities or deficiencies and so forcing their inclusion (Ibid.). If schools really developed an inclusive education, the benefits would be for all the agents involved in the educational process and not just for students with special education needs (Bennett, 1997). Thus, learning communities would be created in order to meet the students' needs with appropriate education support, whether these are immigrants, special education students or students with learning difficulties.

Moreover, teachers' resources and efforts would aim to assess education needs, such as adapting the teaching-learning processes and providing the necessary support to students who request it. Note that in inclusive schools all students are in the regular classroom during the school day; they do not leave and go to a support education classroom. Rather, they get support and individualized attention in the same classroom (Slavin, Leavey & Madden, 1984). The third advantage that Stainback and Stainback (2001) list refers to the possibility of providing social and educational support to all students as far as inclusive school provides support, and promotes student's independence, mutual respect and responsibility. Faced with the definitions that have extended on the inclusive school (Ardanaz, 2004, Gartner and Lipsky, 1987; Stainback and Stainback, 1990, 1992) we must rather define in what respects it differs from traditional integration specially developed in public schools. In this sense, Carrión (2001) states that one must start from the conception commonly shared by the different education agents.

The micro-political level will condition the development that the institution undertakes. It is not likely to identify issues subject to differentiation between the practice of school integration and the development of the inclusive school. Rather, the fact that the factors which show the development of this new approach depend on the configuration of the micro-political level, beliefs and perceptions that school agents share. What issues differentiate school integration from inclusion?

Basically, the term integration refers to the educational response that the school provides to pupils with special education needs. That is, students with learning disabilities, aggressive behavior, mental disabilities, sensory impairments or

physical disabilities (Carrión, 2001: 53). In fact, the use of the term *special education needs* often specifically excludes other pupils, for example, those who were socially, culturally, economically disadvantaged, who had different ethnic backgrounds, etc. In other words—those who are referred to as *late incoming students* by the current Organic Act of Education (2006). A key aspect of inclusion is the individual and tailored attention provided to students with special education needs, whatever their situation is, and especially to those who are disadvantaged, regardless of their origin. This development undoubtedly represents a forward step in the treatment of students who are in social vulnerability.

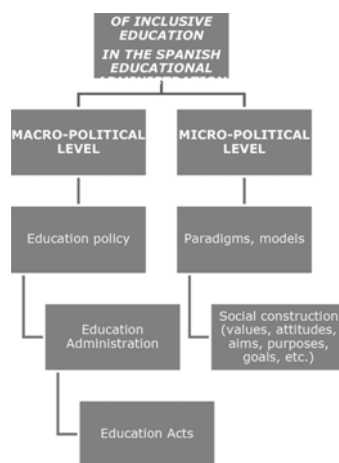
Moreover, it states a redefinition of the term *special education needs*, parallel to the definition of the expression *special education needs* of the Warnock Report (1978), though substantially different regarding its development. To get real inclusion, the education services must provide the attention that each student requires without causing trauma. As Carrión suggests (2001: 59) “avoiding the fact of having negative consequences in those systems who are not able to develop the change it implies”. What changes can be introduced in inclusive schools to develop a truly inclusive inclusion to students with special education needs? What do stakeholders propose to face diversity? What values and attitudes must be shared to reach a genuine education? Firstly, it is essential to use adequate human resources, for example, teachers who really believe in the inclusion of students. They may also have received specific pedagogic training.

Secondly, the use of education materials tailored to the students’ needs. Thirdly, the implementation of appropriate methodology, such as active and participatory approaches, that prompts students into an active role and so on. Fourthly, we must not forget the involvement of parents in their children’s education processes, encouraging an active collaboration, supporting the teacher’s work and offering advice. These assumptions for a deep development of inclusive education should be based on school consensus.

From the macro-political level the central government in Spain (the Ministry of Education, Culture and Sports¹ located in Madrid) through their legislators create education acts. Then the Autonomous Communities, due to the education competences they have, adapt national acts to the reality of each community. For instance, some of them introduce a co-official language, as Catalonia or the Basque county. At this point an Autonomous Community,

¹ <http://debateeducativo.mec.es/paginas/convocatorias.html> (Accessed: 17th November 2015).

through its Regional Government (Provincial Service in the case of Zaragoza) includes in the approved curriculum changes or adds issues to the national guidelines. Education inspectors, consultants, those responsible for the development of education programs, stakeholders, should share values, attitudes and common purposes (Avramidis & Norwich, 2000). When this does not happen the resources to face diversity are ineffective. Many programmes can exist, such those listed below, but they may still be non-practical for inclusive schools if they do not focus on students’ needs. Legislative Acts implement specific measures and resources to develop inclusion but they are no longer developed at schools. When the stakeholders focus on the achievement of those measures, students’ learning highly improves. They are made legal through the adoption of autonomous curricula but lack any efficiency.



The curricula developed in the different Autonomous Communities in Spain are those which the different Elementary and High Schools assume in their didactic projects. As they are prescribed by each Autonomous Administration, the teaching institutions must meet the principles, contents, methodology and assessment criteria stated in those regulations. However, if the arguments exposed in those texts do not take into account the improvement of human and material resources provided to schools, teachers’ efforts towards the development of inclusive education to face students’ diversity will be fruitless. On the other hand, the teacher’s work in the classroom, as mentioned above, depends not only on his/her didactic performance, but also on the methodological line assumed by the school. If the teaching staff is not able to develop the principles, values and aims proposed by the stakeholders, if they do not share a common ground to develop their teaching-learning, teaching is not directed

towards the same goal: to provide each student the support he/she requires. In this sense, we cannot state that a real inclusive education is reached even though the legislative acts implement it. Most of the time what is prescribed by legislation is far from what schools need or what can be assumed in the institution.

Moreover, there is another issue that has to be considered. Multiculturalism is one of the characteristics of our present-day society. As such, schools have a huge variety of students coming from very different countries. Thus, teachers need specific tools and resources. The school, as an integral part of society, must face these new socio-educational demands. The need to provide each student with the attention he/she requires is a constant challenge for teachers. Diversity can be manifested in different orders: unflattering family situations for the student's personal balance, mental or physical disabilities or diseases, interests, preferences and learning rates, etc. Among others, these factors differentiate students even though they belong to the same class group. In order to cope with this heterogeneity, teachers have to undertake specific methodological principles that are not always established by the stakeholders but need to be developed in the micro-political level as those in charge of education are not especially concerned with them.

When a foreign student is schooled he/she has to share his/her values with his/her peers, his/her culture, interest, knowledge with the aim of getting integrated in the group. Even his/her abilities could be different from others. But it will be one of the teacher's tasks to try to encourage an appropriate social environment in the classroom. If the school system does not know how to deal with these socio-educational demands, it can no longer provide quality in education. To make a proper inclusion in mainstream schools, curricular, pedagogical, didactic changes are required. This means that organisation changes must be generated in accordance with the real possibilities and current needs. When this does not happen, if there is not any suitable education management from the school, inclusion can become a special education subsystem that only shares certain organizational and space arrangements with the regular education (Carrión, 2001).

The measures of attention to diversity that the current legislation implements through different stakeholders (inspectors, education policy makers, school advisors, etc.) are meaningless if limited to fragmentary interventions both at the Elementary and High education levels. The inclusion of students with specific educational support needs to be shared at school, in the shaping of the micro-political level. If it is only prescribed by

the macro-political level and even not considered, schools are unprotected towards the new incoming students. Teachers should share common policy principles, goals to be addressed, and also methodological guidelines related to students' diversity. On the other hand, parents have to participate and support the educational process. Education policy should encourage this process of inclusion, not only as keywords that govern the legislation texts, but by providing human and material resources. Many schools need support from administration, not only in terms of staff but also by considering external resources in order to support the teaching-learning processes.

Schools are mainly characterised by a diverse educational reality which is quite difficult to be assumed. They do not only need more material resources, but often human resources are not sufficient or the most adequate ones in such an increasingly diverse and multicultural environment as the current one. In this sense, schools have to undertake common principles to effectively manage the centre through a coherent and coordinated action of all members of the education community. This view must be shared by stakeholders in order to support education in each stage. In Spain, specifically in Aragón, the assumption of these common principles, values and attitudes is usually reached in schools to offer the education required by each student. However, even though sometimes they are stated in the legislative text and are orally extended by the stakeholders, they use to remain as perfect words, legitimate words teachers must use in their speeches. As an example of how stakeholders' indications are somehow reflected in official education documents we necessarily have to mention the School Educational Project. If designed appropriately and coherently on the basis of the schools' needs and the stakeholders' prescriptions, it will favour the development of an appropriate school inclusion.

The School Education Project is a document that shows how the school conceives education, what values, aims and structure it has and how education is developed. It is created in schools following the stakeholders' prescriptions. If the information reflected in this document were truly assumed in schools, we would state that the stakeholders' guidelines were efficient and functional at schools as far as they were developed. Unfortunately, this document does not always show the real development of the educational process. Even though the main aim of the guidelines stated by the stakeholders is to improve education, by offering the right answer to all students, the macro-political level is not always related to the micro-political level.

To give an example about how this document shows a coherent development of the education

assumed by the school staff, and prescribed by stakeholders, we should mention the three fields it focuses on: a) the educational field (how the school assumes the education process, what values, attitudes and purposes the school develops); b) the institutional field (school management, organization, resources, etc.); c) the administrative field (teachers, specialists in Therapeutic Pedagogy, number of students per class, timetable, etc.). The norms to create this official document are stated by the legislation implemented by the stakeholders in the macro-political level. Furthermore, the development of this document has to deal with the micro-political level as far as it is going to be developed in a specific school. First of all, if we focus on the educational field, we may pay attention to the school organization, cycle equipment, level staff, specific methodology, the assumption of common criteria for students' promotion, students' grouping, evaluation, etc. Secondly, the institutional field is conditioned by the school external relationships (relationship with the Parents Association, Administration, etc.) and the internal relationship in the school (government/management bodies, functions they assume, commissions, degrees of participation, etc.). Third and finally, the administrative field related to school economic management, accountability, uses of space, mass-media rooms, etc. Human resources and interpersonal relationships (motivation, communication, and conflicts), regulation of coexistence, selection and promotion of teachers, etc. should never be overlooked. These aspects and the actions involved in education policy hinder the effective management of the school. The school needs tools to support the actions of the educational community. The teacher that faces multiculturalism and students' diversity acts as a facilitator of the teaching-learning processes. Thus, the development of the educational practice starts from the requirements of prevention and the settlement of students' difficulties, understanding that both conditions must be implemented as soon as possible in the school environment.

The attention to classroom diversity must be assumed in terms of cognitive abilities and in terms of cultural backgrounds. It is therefore imperative to understand that differences and inequalities should be addressed through affirmative strategies and not by universal or standardised solutions within the framework of an inclusive school. This implies the assumption of the principles of intercultural education in line with the growing multiculturalism and diversity that characterizes our society. These actions will favour citizenship education to let students acquire an integral education, a life-long learning process. Inclusive education should be built in an educational community as a general action framework integrated by all

agents involved in the education process. A cooperative endeavour of all education agents and every moment of the school life is essential. In order to achieve these aims, we should promote the students' interest in school work, by reinforcing the previous motivational processes. Moreover, it is also necessary to support the students' learning processes by seeking individual attention, without affecting the group dimension in the classroom work.

Similarly, it is advisable to keep an adequate environment to favour coexistence in the classroom and in the whole school as a prerequisite for the appropriate development of the school work. We do not have to forget that it is necessary to help to reduce school absenteeism and the subsequent students' failure. We should also strengthen cooperation with families, other neighbourhood institutions and other social settings. Therefore, inclusive education in school needs to establish links that extend and enhance the educational efforts beyond the school environment. Finally, it is also convenient to enhance the process of education innovation, further evaluation and investigation that should feed all innovative proposals. If schools contribute to the development of inclusive education, the existing measures to address students with specific education needs increase as long as the learning processes are based on their interests and this creates motivation in students. These principles not only sustain the so-called "inclusive school" but also favour integration and inclusion of disabled students.

3 DIDACTIC STRATEGIES TO APPROACH SCHOOL INCLUSION IN THE SPANISH EDUCATION SYSTEM

Once the stakeholders of inclusive education have been identified, we should analyse how teachers implement the educational processes when dealing with diversity. Obviously, the principles, values and objectives that are reflected in the School Educational Project must be coherent with the educational line of the Administration. It is not enough to establish pattern behaviours but it is a much more complex matter. To be able to state that a school really develops inclusive education, the school has to combine in its Educational Project and its practice of the education policies, established by the stakeholders and the creation of a school community, concerned with how to deal with multiculturalism, diversity and heterogeneity. This is only possible if teachers build the school reality from their daily interactions. In this way they share a culture which will undergo changes over time as actors or agents also shift. The school, therefore, is a social construction of its members

but it is conditioned by the ideological pressures that come from the macro-political level. These pressures are generated by the stakeholders who implement new education activities, methodologies and principles reflected in the legislation documents. But the common culture shared in a school leads to share values, attitudes and aims that must be related, in turn, and be compatible with those prescribed by the stakeholders. Otherwise, these actions would make no sense.

In the high school where this case study has been developed, an exhaustive and systematic observation to detect the school's culture has been conducted. This observation has allowed us to affirm that inclusive education, with all the issues it implies (values, principles of the didactic process, methodology, tutoring, etc.), has been implemented in a correct way. Not only by what is prescribed by stakeholders, but, rather, thanks to the collaboration of all staff in the assumption of a set of values that govern the teaching-learning processes and that are made explicit in their School Educational Project. The school methodology focuses on the improvement of a way of learning based on realism, close to the students' foci of interest in order to increase their motivation and promote their autonomy in the development of school tasks as well as in the learning of educational values.

First of all, it is essential that both the school and the environment, and the social reality that it belongs to, are in a close and constant interaction (life-long learning, functional learning...) (Canals, 2004). In this way, the teacher has to know, seek and properly use the possibilities offered by the environment, either in educational, social, cultural terms, tasks, resources, etc. This aspect is essential in the educational process as it determines the effectiveness of our action, and may increase students' motivation, by improving educational services for incoming students or completing the teaching-learning processes in real contexts through meaningful learning.

Second, the didactic process must be based on globalization as an essential methodological principle, as far as it is a mandatory enrolment period, and so far as it is prescribed by current legislation. The contents that the student with specific needs of educational support works on must be closely interrelated with other curricular areas. This means creating educational materials and activities that proceed from the student's reality and which are developed through interdisciplinary projects. In this way, his/her motivation will increase as well as his/her perception of integration in the class group. Therefore, the tasks have to move towards interdisciplinarity, being developed through workshops, specific programs, stories, role-plays, etc.

Third, closely related to the principle of globalization, we highlight cooperative learning. Since the

transmission of knowledge is conditioned by the students' diversity, it is necessary to promote the student's socialization. The social dimension of learning is implemented through cooperative learning, since it is a construction that we do not develop alone but interacting with others. Therefore, it has a social component. The teacher, from the interpretation of his/her classroom reality, diversity and heterogeneity, from his/her background, will decide what to do -mediational paradigm centered on the teacher (Pérez, 1983)-. He/she is therefore a "reflective planner" (Ibid.: 118) of the teaching-learning process. He/she abandons the standard models -typical of the traditional process-product paradigm- and understands that the teaching-learning processes have to be planned ahead with the aim to develop an initial assessment and provide students with adequate individualized attention.

Cooperative learning involves the development of cooperative skills in students that encourage and improve communication styles, which are essential especially in the disabled students in order to enhance their socialization in their group-class. This learning methodology also involves the use of language as an instrument of dialogue and communication, by building confidence in students when it comes to expressing their opinions and minimizing the consequences that may pose conflict.

Fourth, so as to be cooperative, learning has to be based on a well-done classroom space distribution, according to the characteristics of the task and using the resources required. Therefore, the allocation of roles, resources and space will encourage the development of cooperative skills as well as communication styles. In fact, the direct correlation that develops between the tasks of instruction and classroom management (Doménech, Traver, Odet & Sales, 2006) brings new ways in the use of classroom spaces. The homogeneous or heterogeneous students' grouping either determined by teacher's selection or through the teacher's instructions, conditions the interaction between students, their cooperation in learning and their socialization. The classroom is a social space for exchanging experiences, culture, autonomy, communication and socialization (Ibid.). Thus the coexistence of students' heterogeneity and diversity must be improved. In this sense, this social space ought to promote active participation of all students, interaction, exchange, dialogue, etc. Attitudes of acceptance or rejection may arise but it will be a teacher's task to ensure that these are transformed into bonds, self-knowledge and mutual learning. Thus, the development of guided learning processes will determine the classroom social environment (Arnáiz, 2003). This constitutes the fifth methodological principle, as it is essential in the establishment of techniques, principles and strategies in the

teaching-learning processes. The social classroom environment is an essential element in the establishment of interactions and cooperation among students. While a number of factors are involved in its configuration (such as the political-economic-administrative agents, social relationships, communication, culture and architectural features, temporal sites, etc.), its marked complexity provokes the generation of new instruments of mediation. These instruments follow models of social relationships dependent on instructional tasks, on the teacher's and student's roles, as well as in the classroom setting itself (Slavin, 1986).

These relationships established in the classroom can be of cooperation, competitiveness, self-reliance, empathy, rejection, activity-passivity or equality-inequality (Ibid.). Proposed in a dichotomic way, they always set undoubtedly the prevailing culture of the institutions, its beliefs and interpretations of learning. Therefore, the classroom is not an enclosed space, but it is open to dialogue and interaction, in which knowledge is built from experience, reality and students' mediation-intervention. In this sense, the teacher must foster a flexible social classroom environment, open to students' diversity, interests, abilities, skills and attitudes. In other words, it must promote interaction and cooperation among students as an essential element for learning.

The group distribution in the creation of the social classroom environment to develop specific tasks will condition the social relationships among the students and thus promote the inclusion or not of disabled students. If learning is the acquisition of new knowledge from the establishment of social relationships, educational experiences that are generated in that social classroom environment are essential for the acquisition of concepts and the integration of students in society and ultimately in public life (Gimeno, 2008).

Sixth, the teacher should promote the principle of activity. That is, the student has to be the main protagonist of his/her own learning, which appeals to his/her intense cognitive activity, more or less accompanied by motor or manipulative skills and always facilitated by a logic gradation of complexity in tasks. The student becomes an active element in the teaching-learning process. This implies that the teacher does not fulfil the role of transmitting knowledge; s/he rather acts as a facilitator of the student's learning (Montero, 1991). The student then learns to learn. This methodological principle, especially widespread in current times with the introduction of the eight basic competences in the Spanish context, allows students to learn and build knowledge with the teacher's input. Thus, independent learning is promoted. If our social classroom environment is adequate, and disabled

students feel included in the class-group, they can share and build knowledge together. However, the teacher should encourage student's activity. If not, there is no learning. If we want a stimulating and attractive task, we have to focus on the student's centres of interest, on his/her reality. Therefore, the teacher must find out the students' concerns and build new relationships with the contents of the curriculum. To promote the principle of activity in the student, to provide his/her with an active role in the development of tasks, implies to select those that arouse more motivation, which are tailored to his/her skills, cognitive level and addressing towards self-learning (Perret-Clermont, 1988). Tasks involving classroom research, experiments, interaction, debates, brainstorming are useful in this regard.

Seventh, students must reach a meaningful learning. He/she has to be able to integrate his/her prior knowledge to the new acquisition, so that the latter will be meaningful, both in the receptive and constructivist learning. Only if this condition is reached, can we talk about functional learning and enable the principle of learning to learn: "The main education paradigm to be followed should be to learn to learn and learning by doing". Meaningful learning is necessary to transfer knowledge to real life situations which involve problem-solving tasks" (Pérez, 2008: 136).

Eighth, the principle of individualisation must also be considered. In order to provide students with an individualized attention we may start from their level of cognitive competence, their previous knowledge, pace of work, etc. (zone of proximal development, Vygotsky, 1978). This involves a system of reinforcements, rewards, incentives... to the achievement of the learning aims that must be immediate and very tight to their interests. The teaching-learning processes must be adapted to the characteristics of each student, as they are different and have different cognitive processing. Similarly, their interests, motivations and concerns are far from each other; thus teaching practice should include an individualized instruction tailored to each student's needs. The cards, individual work, concept of mapping, etc. are of great help in the process of individualized instruction. It is through the implementation of cooperative learning where individualization is made compatible with this methodological principle, by diversifying the teaching-learning processes. These kinds of tasks improve both the socialization and inclusion of disabled and immigrant students and the acquisition of learning.

These ten methodological principles must be developed by using different teaching resources, different means involving different languages (auditory, visual, audio-visual, body language, etc.).

In this sense, we stimulate all students' senses, not just language. Especially relevant are Information and Communication Technologies (ICT) in this education context. The computer, the digital interactive whiteboard are very motivating for these students, by allowing them to individualise their learning pace. Moreover, classroom organization is especially relevant in the processes of school inclusion. Students with special education needs must feel comfortable. The classroom distribution may favour communication between the teacher and the students. There are different ways to group students in the classroom. The most suitable are essentially flexible students' grouping formed by appointment of the teacher, students' free choice, or with heterogeneous groups (always keeping in mind the gender) or homogeneous in terms of learning levels.

Groups can develop the same task or carry out differentiated tasks, which are then pooled. In this sense, work in pairs is also an option. Obviously, given the students' characteristics, the teacher may then establish working groups to be more effective for the purposes of socialization because they foster close relationships between their members. However, depending on the aims to be pursued in each teaching situation and the characteristics of the task, a clever combination of all the options outlined above seems to be a most suitable strategy to keep a social classroom environment. However, besides organizing the classroom, the teacher should consider organisational criteria, such as teamwork and cooperative learning and coordination with other collaborating agents (coordination of teachers with families and professionals involved in the education process).

To face students' diversity in any school, horizontal organization must favour the relationship between and among different forms of learning, by providing students with strategies to train themselves into adulthood, to let them acquire a social inclusion and reduce their risk of exclusion. The United Nations Convention on the Rights of Children (1959), in the Article 28, established the right of all children to get a basic school education based on equality of opportunities, regardless of their cultural, social or ethnic backgrounds.

The current proposal aimed to promote the education of students with specific education needs should be based on continuous learning, on life-long learning, on cooperative learning. Thus it has to focus on respect and acceptance of individual differences. This involves the development of the principles of inclusive education stated in the School Education Project through the active involvement of the staff, through the human, material and technical resources required, and through the support of the stakeholders. This educational action must promote the teaching-learning

processes in a diverse context, sensitive to multiculturalism present in our society and reflected in the didactic intervention on the basis of the principles of multiculturalism. This project builds on the school culture and the values it assumes. It would be appropriate that these values, principles and aims were also undertaken by those raised by the stakeholders in the legislative acts.

The so-called schools of difficult performance are still a challenge for teachers. The principles of inclusive education contribute to the decline in the number of illiterate population. They provide a solid foundation in the establishment of inclusive education, adapted to the students' needs, their difficulties and being compensatory to reach equal opportunities in education. It is not enough to provide the conditions to make smaller groups in the classroom, or to obtain support from specialist teachers in Therapeutic Pedagogy and Special Education, as stated by the stakeholders of macro-political level.

4 CONCLUSION

Students' diversity and heterogeneity in Spain is endless, especially in Elementary and High Schools. Indeed, it requires highly skilled and motivated professionals to be able to meet that diversity. The specific pedagogical processes must contribute to the creation of a social classroom environment where the student feels comfortable enough and motivated towards learning. Firstly, the stakeholders determine the most relevant aims, basic competences, contents, methodological principles and assessment criteria to be taught and reached at Elementary and High schools. As far as these are prescriptive, all schools nationwide must develop them. Subsequently, the different Autonomous Communities in Spain adapt those prescriptions to their needs.

In the case of Aragon, the Education Administration provides more resources to face students' diversity. However, this lack of coherent development from the macro-political to the micro-political level keeps a handicap in the attention provided, especially to students with specific need of education support. The main aim is to offer students the individualized attention they require to be able to develop their cognitive and social skills. The schools, on the other hand, set their aims and priorities, following obviously the prescriptions of the Autonomous Education Administration. Among them, learning to learn, being adult, responsible people and acting as such are key principles in any compulsory educational stage. Other basic aims deal with the development of inclusive education, the improvement of students' achievements, the integration of disabled students, etc. The principles

that the school assumes, through the sharing of a common culture built up through daily interaction of all staff members, are reflected in the aims pursued by the school, as they are reflected by the School Educational Project.

However, the methodological principles that help in the development of inclusive school are far from the prescriptions and interests of the stakeholders, as the latter do not know the current needs of school. They are not teachers in any classroom, and everything they propose is more theoretical than practical. Secondly, the teacher's work does not only focus on the students' assessment. He/she has to plan his/her teaching-learning processes create a syllabus design, consider the students' diversity to propose specific measures to deal with that heterogeneity, etc. When planning the didactic processes, we must do it permanently throughout the learning process (continuous assessment). In this sense, we get the necessary feedback to introduce any changes that may be considered. In other words, teaching is "a decision-making" process, where the teacher is a permanent "decision maker" (Pérez, 1983: 116–117).

When we develop a syllabus design, we decide what to do in the classroom, on the basis of considering the students' characteristics. When carrying out our didactic programming, and depending on the incoming needs, we will consider whether we continue with those aims or we change them (formative assessment). This decision must be taken on the fly, while we develop our teaching-learning processes in the classroom, when we finish the session, the didactic unit, etc. Thus we can assess the learning achieved (final assessment) and once again we get new feedback from our programming for the next didactic process. Certainly, the learning process of each student is different and so is the result. The knowledge that each one conquers is his/her cognitive construction, peculiar, possibly unique. The student is the mediator for excellence of his/her own learning, the main protagonist. While he/she is learning, a set of interests come into play, closely related to his/her previous life experiences.

The knowledge he/she already possesses, and also the cognitive processes that enables him/her to activate them, depends on the input he/she receives. This is the fundamental premise of the mediational student-centered paradigm, in line with the approaches of the cognitive psychology, which understands the teaching-learning praxis as a process that should facilitate the construction of knowledge and the development of student's processing information strategies (Pérez, 1983: 120–122), both in his/her individual work, and in cooperation with his/her peers, and under the teacher's guidance and support.

Thirdly, schools that develop an inclusive education should refine the concept of teaching and

learning and attune it to the real achievement of the students' priority aims. Only then can we say that the implementation of an innovative methodology in the treatment of compensatory education contributes to overcoming inequalities in education, to spread schooling to disadvantaged, marginal contexts, and to deal with the growing multiculturalism of the classroom. However, other actions to be undertaken by the whole of the school community could be considered to help in this process of education intervention. For example, the creation of discussion groups, composed by staff members, and in other cases representatives of parents and other education agents, would feed a situation of permanent critical analysis of the school educational development process. The internal and external assessment of the School Educational Project would complement the performance of different education agents with the input of external agents, experts in education innovation, with the participation in dissemination and discussion forums on education innovations (with gypsy students, immigrants, disabled...). Those discussion forums could be developed both in school and online. Finally, the participation of teachers in life-long learning processes to improve their teaching skills, get more strategies when dealing with diversity, etc.

Moreover, the organisation of life in the classroom in such schools is especially relevant. It requires the assumption of methodological principles set out in the School Educational Project, in order to promote students' interaction, equality of opportunities and non-discrimination.

Teachers must establish flexible student groupings, raised within the class group or groups, involving two different groups of the same cycle. Strictly speaking, a non-graduate teaching within each cycle could be implemented, provided that the non-graduation work would focus on the development of instrumental techniques. Depending on the aims of each didactic task, both strategies could be used. Finally, flexibility in planning and the use of space and time must be open: flexibility in students' grouping requires acting in accordance with the classroom space distribution or cycle spaces and, in turn, teachers have to distribute time depending on the activities to be developed. As a general rule, short time periods are convenient for each subject, task, in order to avoid fatigue and students' disinterest.

REFERENCES

- American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders (DSM-IV)* (4th ed.). Washington: Washington, DC.
- Arceneaux, L. S. (1993). *The influence of Teacher Behaviour on the Distribution of Achievement in the*

- Classrooms: An Application of the Hierarchical Linear Model*. Doctoral Dissertation. Baton Rouge, Louisiana: Louisiana State University.
- Asher C. & Malet R. (1996). The IUFM and initial teacher training in France: socio-political issues and the cultural divide, *Journal of Education for Teaching*, 22(3), 271–281.
- Avramidis, E. & Norwich, B. (2000). Teacher's Attitudes toward Integration/Inclusion. A Review of the Literature, *European Journal of Special Needs Education*, 2 (17), 129–148.
- Bragg, L. (1997). From mute god to the lesser god: Disability in medieval Celtic and Old Norse literature, *Disability & Society*, 12, 165–177.
- Carrión, J. J. (2001). *Integración escolar: ¿plataforma para la escuela inclusiva?* Málaga: Aljibe.
- Doménech, F.; Traver, J. A.; Odet, M. & Sales, M. A. (2006). Análisis de las variables mediadoras entre las concepciones educativas del profesor de secundaria y su conducta docente. *Revista de Educación*, 340, 473–492.
- ETF - European Training Foundation (1997). *The VET System in Albania-Recent Changes, Challenges and Reform Needs*. Tirana: Albanian National Observatory Institute of Labour & Social Affairs.
- Gallaudet, H. (1998). *Working Papers 89-3*. Washington, D.C.: Gallaudet University.
- Goodman, N. (1989). Education for critical democracy. *Journal of Education*, 171, 2, 88–115.
- Howe, S. G. (1866). *On the proper role of state institutions for the disabled. Speech given at ceremonies on laying the cornerstone of the New York State Institution for the Blind at Batavia, Genesee County*. New York: Henry Todd.
- Jurado, P. & Soler, R. (2015). Workers with disabilities in sheltered employment centres: a training needs analysis, *International Journal of Inclusive Education*, DOI: 10.1080/13603116.2015.1111446.
- Organic Act 2/2006, of May 3rd, of Education (2006). In Boletín Oficial del Estado, 106, 17158–17207.
- Gimeno, J. y Pérez, A. I. (2008). *Comprender y transformar la enseñanza*. España, Madrid: Morata.
- Perret-Clermont, A.N. (1988) (Ed.). *Interagir e connaître: Enjeux et régulations sociales dans le développement cognitif*. Neuchâtel: Delachaux et Niestlé.
- Soler, J.R. (2009). La participación social en la construcción de la democracia, reto consustancial a la formación a lo largo de la vida. *Libro de Actas del V Congreso Internacional de Formación para el Trabajo*, 399–413. Madrid: Editorial Tornapunta Ediciones.
- Soler, J.R. (2013). Estado actual y estrategias para futuros de la formación a lo largo de la vida. *Proceedings of VI Congreso Internacional de Formación para el Trabajo*, 369–379. Zaragoza: Editorial Tornapunta Ediciones.
- Soler, R. (2012). Is the teacher's discourse creative? Analysis of its most frequent expressions. *Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 10(3), 88–104.
- Soler, R. (2013). Acciones educativas para colectivos en situación de vulnerabilidad social en Aragón: Alumnos inmigrantes con diversidad lingüística. ¿Cuál sigue siendo el problema de fondo? *Educar*, 49(2), 267–286.
- Soler, R. (2014). *El poder de las palabras: un análisis del lenguaje pedagógico*. Zaragoza: Mira Editores.
- Soler, R. (2015). Lenguaje y práctica educativa. Claves de la terminología pedagógica. Dykinson, Madrid.
- Zamorski B. (2006). *Bringing Industry and Academia Closer Together: The Introduction of the Foundation Degree in the UK*. In P. Tynjälä, J. Välimaa & G. Boulton-Lewis (Eds.), *Higher Education and Working Life—Collaborations, Confrontations and Challenges*, Oxford and Amsterdam: Elsevier, 57–72.

An empirical study of the MIS impacts and organizational efforts on improving business performance

Azza Abdel Monem

Faculty of Computer and Information Sciences, Ain Shams University, Abbassia, Cairo, Egypt

ABSTRACT: The Management Information System (MIS) has become an integral part for running and managing a successful business today. In many industries, survival and the ability to achieve strategic business goals are difficult without extensive use of information systems. The information system represents a combination of management, organization, and technology elements which provides solutions to problems or challenges facing a firm. This paper focuses on the impact of inappropriate and unstructured management information systems and organizations on the firm performance. The study identifies some problems in Sony Pictures Entertainment Company and suggests practical solutions to improve the efficiency and effectiveness in the company's daily activities by adopting a structured MIS. The main finding of this paper will help to understand the importance of MIS in assessing, scheming, applying, implementing and exploiting technology in producing information in order to improve the performance of decision making, including decision support systems, expert systems, and executive information systems.

1 INTRODUCTION

MIS refers to the study of organizations, people, and technology, and how the three elements are interrelated or interconnected. Information and technology are the two most important components of any organization. It is common organization practice that the manager has to ensure that the effectiveness of information systems is optimized in order to enhance the overall levels of efficiency and organizational output. The role played by MISs within an organization does not only revolve around the implications of technology but also on the performance of employees (Grant, 2010). For instance, it is an excellent platform that serves as a mechanism for decision-making.

Over the years, different companies have employed distinctive approaches for the implementation or integration of MISs within their respective operational frameworks (Sungjune 2015). While some have effectively implemented MIS, others have encountered numerous problems (Alec Cram, 2016). This accentuates the need to effectively analyze the distinctive organizational needs prior to the implementation of MIS. Additionally, the relevant technology should be employed in the enhancement of the overall effectiveness of MIS within organizations (Kayworth, 2008). Based on these stipulations, it is evident that there are numerous factors that contribute towards the overall implications of MIS within an organization. This analysis zeroes in on the importance of MIS.

A case study of an organization that has effectively implemented MIS within its operational blueprint will also be evaluated.

2 RELATED WORK

MISs are characterized by extensive implications on distinctive organizations where it has been implemented. Different researchers have evaluated the importance of MIS from different angles as elucidated in this section.

2.1 Cost reduction

One of the most pertinent aspects that exemplify the importance of MIS is the reduction of operational costs. Cost efficiency is an integral element of any organization. Companies cannot attain their stipulated targets if they do not have relevant frameworks for the mitigation of costs (Gibson, 2010). On the other hand, the operational efficiency of companies in the business world is immensely enhanced when the framework of cost reduction is implemented. This accentuates the importance of ensuring that high standards of cost efficiency are attained in an organization, including MIS implementation. Numerous researchers have evaluated the distinctive ways in which MIS enhance the framework of cost efficiency.

Firstly, it is notable that MIS reduces the space required to store the data because MIS is grounded

on modern digital systems. The overall physical space required in the maintenance and storage of data is reduced through the implementation of MIS (Haag, 2012). This is a pertinent aspect that contributes massively towards the overall framework of cost reduction in an organization.

Another dimension or way in which MIS contributes towards the reduction of costs is the issue of time management. In the absence of an effective framework of MIS, an organization is strongly likely to encounter extensive challenges with regard to the time it takes to transfer information across different employees or among different employees. So, MIS contribute towards the reduction of the time taken to complete tasks (Gupta, 2010).

MIS serves as a pertinent platform on which the relevant mechanism of research and development are implemented in an organization (Kennedy, 2005). One of the most pertinent aspects of consideration is the identification of the best ways in which the overall costs or expenses of an organization can be reduced. This is why MIS serves as a vital aspect with regard to the overall blueprint of costs reduction.

2.2 *Innovation*

Another outstanding role of MIS within an organization pertains to the enhancement of innovation standards. In the 21st Century, innovativeness is an integral element of the strategic blueprint of most organizations. This is because of the inherent ability of innovativeness to promote internal productivity, the quality of outputs, and overall levels of competitiveness, among others.

The importance or role of MIS towards organizational innovativeness can be evaluated in numerous ways. Firstly, MIS facilitates for the alignment of knowledge systems with the available resources (Ashlyn, 2006). Such an alignment is integral towards the attainment of new strategies for doing business or production.

Additionally, collaborations between different companies are vitally essential in terms of cost reduction (Haag, 2012). Innovation is pertinent because it facilitates for the expansion of the business framework of an organization. It also serves as an influential element that plays a massive role towards the enhancement of the ability of organizations to deliver on their mandate.

2.3 *Enhancement of marketing efficiency*

The enhancement of marketing efficiency serves as another crucial implication that accentuates the importance of MIS in an organization. Almost all organizational goals are directly connected to the attainment of marketing efficiency. This is

due to the fact that the absence of such systems would negatively affect overall performance of the organization (Watson, 2015). The world's best performing companies in the global business sector are always associated with exceptional frameworks for marketing.

The core of any marketing system is information and how it is harnessed. This underlines the fact that MIS is an important aspect that plays an influential role towards any marketing system. Firstly, it ensures that the marketing team of an organization makes informed decisions as pertains to the choice of marketing strategies. While some marketing techniques are effective in the promotion of certain products, they are more or less ineffectual when it comes to the promotion of other products. Additionally, the strategy used in the promotion of a product to a given target group might be ineffectual as a result of demographic dynamics (Effy, 2008).

The importance of MISs with regard to marketing can also be evaluated in terms of the contribution it makes towards customer relationship management. For example, electronic CRM has become increasingly popular because of the part it plays in the enhancement of customer service through modern technological systems. This enhances the swiftness with which the inquiries of customers are addressed. This is vital in marketing because it helps in the promotion of a company's brand (Laudon, 2014).

2.4 *Control of creation, retention and growth of records*

The concept of management of any organization records can virtually become meaningless if the growth of records is not taken care of (Gupta, 2010). For years, papers are still increasing in various offices across various organizations. Reportedly, such increase of paper work has been majorly due to poor management of such records (Gupta, 2010).

Eventually, such piles become complex enough in the sense that separating still-important files from the unimportant ones becomes a cumbersome process. A properly MIS ensures a way to prevent such scenarios from occurring (Heijden, 2009). This means that there must be a mechanism that helps to determine the need to keep various records in paper form or not to. However, without proper strategy of disposing information, there can be another problem such as spillage of information to unintended audience. This and other cases such as loss of vital data are factors that necessitate any controller of the records to be very keen when handling different pieces of information in an organization.

2.5 *Improving efficiency and productivity*

It goes without saying that properly MIS is a fundamental factor of organizational efficiency and productivity. Efficiency is directly related to productivity. Therefore, efficiency in an organization is of paramount importance (Gupta, 2010). For example, a properly managed information systems ensures that there is minimal time wastage in the day-to-day's operations of within a particular organization or a production process (Effy, 2008).

2.6 *Assimilation of new record management systems*

A good MIS easily provides reliable data that can easily help in adopting new information management technologies. If, for example, an organization is prepared to enter its manual data into a computerized system, such a process can be an easy hit if the manual data is analyzed beforehand (Shajahan, 2007).

Technically, an organization is able to accrue the benefits of new technologies that otherwise keep entering the market. Keeping information systems organized and data analyzed as needed at all times always places particular organization at a very advantageous point in the current age of rapid growth in technology (Gupta, 2010).

2.7 *Ensuring regulatory compliance*

In many operational regions, regulations vary greatly. Depending on the kind of regulations operational at a particular place, there is no other best alternative for an organization to ensure compliance than referring to its previous records or else working with proper MIS. A proper MIS keeps track of day-to-day activities of an organization and therefore automatically ensuring that such an organization complies with the set regulations.

2.8 *Helps in minimizing litigation risks*

Proper MIS help to keep the organization in question safe from possible litigation risks that can end up in imposition of hefty penalties (Heijden, 2009). Various litigations may occur in forms of complains against infringement of a particular stakeholder's rights, litigations emanating from confusion of roles in the work place, loss of corresponding government or regulatory compliance records of past periods of time, and so on. Research has shown that there are many organizations that have undergone litigation processes involving heavy compensation imposition due to just loss of records (Heijden, 2009). Proper record-keeping keeps such litigation threats away and also

keeps reduces workplace conflicts (Effy, 2008). This shows that proper MISs serve as a cushion against ill intentions, politicized litigations, corruption allegations and other possible atrocities such as copyright infringement and so on.

2.9 *Protection of rights to intellectual property*

Aforementioned above, the capacity of a MIS to protect an organization from litigation does not end there: protection of rights to intellectual property is specifically a more sensitive entity. In some years back, there ensued a conflict between Microsoft and Apple that involved copyright infringement (Shajahan, 2007). While Apple succeeded in the case, it was only a matter of recordkeeping that solved the case. Microsoft was ordered to compensate Apple. While it is Microsoft that had stolen Apple's copyright, this case was all happening due to spillage of vital organization information into malicious audience (Shajahan, 2007). There are many other ways through which poorly managed MIS can cause disastrous ends to the maiden organization or business entity. For example, lost information can accidentally land into hands of malicious people, mostly competitors, who are more than likely to destroy a particular organization's reputation and/or steal copyright.

One reputation of Apple management is that the company tries as much as possible to keep its copyright and technological achievements well hid away from the competitors (Gupta, 2010). This has with no doubt helped the company to stay afloat and eventually the company is rapidly gaining more percentage of market share compared to other similar organizations. But this achievement has not been easy so far: Apple ensures high wages for its staff such that the staff has no reason to betray the costly trust bestowed on them. Therefore, a combination of human resources management system and information security system has formed a formidable mix that has hitherto kept Apple afloat (Effy, 2008). Others like Sony have suffered malicious ordeals such as system hacking which led to the organizational shaking, all due to lack of proper MIS.

3 CASE STUDY

Sony Pictures Entertainment is one of the world's largest and most profitable corporate organizations. It specializes in film production but it has recently ventured into other business areas. Despite the company's sustained success over the years, it has been characterized by MIS that have been prone to numerous problems. For instance, the company's database systems were hacked in 2014

leading to extensive damage on the organization's reputation (Laudon, 2014). This incident also led to the release of confidential information about Sony Pictures Entertainment. This section focuses on the inappropriate MIS that has characterized this company in recent years.

3.1 *Inappropriate MIS at Sony Pictures Entertainment*

As briefly outlined in the previous section, Sony Pictures Entertainment is one of the world's most reputed organizations in the film production sector. It has been behind the production of some of the world's chart-topping films of recent times. Additionally, this organization is characterized by excellent revenues that translate into exceptional profits. However, the company's MIS have been largely ineffective especially considering the scale of operations. One of the most notable aspects about the inappropriate MIS at the company is that there lacks an effectual framework for database protection. This means that the organization has been immensely exposed to copyrights infringement as a result of the weak MIS (Watson, 2015). Another notable attribute of the MIS in this company is that it has not been aligned effectively towards the distinctive company goals and long-term objectives. This is an attribute that has significantly contributed towards the inappropriate MIS at the company.

3.2 *Implications of inappropriate MIS at Sony Pictures Entertainment*

The absence of appropriate MIS at this company has had numerous implications. As briefly outlined earlier, this organization was rocked by a major hacking incident in 2014. Hackers accessed the database of Sony Pictures Entertainment and made away with confidential data of about 100 terabytes. After the initial assessment of the incident, it was evident that the hackers had been accessing the company's database for a long time (Laudon, 2014). This incident was totally unprecedented because a company of Sony's magnitude is expected to be characterized by a highly effective MIS. Additionally, it is notable that the hacking incident was well-orchestrated and some insiders might have even collaborated secretly with the hackers.

One of the most notable outcomes of this incident was that the reputation of Sony Pictures Entertainment was damaged. The credibility with which the company had been perceived for many years began to vanish. Apart from reputation, it is notable that the brand of the company was damaged severely by the incident. This is due to the

fact that potential customers and business partners began to doubt the ability of Sony to safeguard information and confidential data. The implications of inappropriate MIS at the company can also be evaluated in terms of the manner in which the confidentiality of employees was affected. In essence, the hackers exposed extensive personal information of numerous employees of the company. This included private emails between senior executives at the company. This is a glaring indicator of the extensive havoc that can emanate from the absence of an effective MIS in a company (Watson, 2015).

3.3 *Solutions*

In view of the distinctive negative implications of inappropriate MIS at the company, there is the dire need to evaluate the distinctive strategies that can be implemented in order to avert the problem. Firstly, it is immensely crucial for Sony to invest heavily in the acquisition of modern technological systems from credible manufacturers. This is an approach that would immensely enhance the ease with which the effectiveness of MIS at the company is achieved. Additionally, purchasing such systems from reputed manufacturers would facilitate for the avoidance of MIS that is inconsistent with the distinctive needs or operational requirements of the company. From another perspective, another crucial aspect of consideration with regard to the MIS framework of Sony is the alignment of the entire MIS architecture to the distinctive mission and vision of the company (Clarke, 2007). This is an important element that contributes towards the failure or success of MIS frameworks in any given organization. In view of this stipulation, there is the dire need for Sony to ensure that all MIS technologies are in line with the stipulated goals and long-term strategy of the company. Such consistency is vital in terms of facilitating for uniformity with regard to the approach used by all employees and their attitudes towards MIS. From a different outlook, it is massively critical for the company to recruit highly skilled professionals in MIS.

The absence of presence of such skills is an aspect that contributes massively towards the effectiveness of operations in any given organization as far as the framework of MIS is concerned. When an organization has highly skilled employees, it becomes easier to attain the required standards of implementing MIS. On the other hand, the absence of highly skilled employees undermines the overall framework of implementation. This accentuates why Sony must invest heavily in the recruitment of experts in MIS. Such employees will develop a comprehensive blueprint that will

serve as a reference point for all MS aspects at the company (Bagad, 2009).

This is an approach that has facilitated for the successful implementation of MIS in companies such as Wal-Mart. In line with these implications leaders should employ the best decision-making approaches while identifying the resource requirements for MIS at the company. From another approach, it is also essential for Sony to invest in research and development. This is an approach that enhances the company's ability to monitor its technological infrastructure and also train all employees about the distinctive functions or significance of MIS in the day-to-day operations. Such an approach has been densely effective in other companies, and would hence be vital for integration at Sony Pictures Entertainment.

Additionally, monitoring is pertinent in terms of aligning the entire framework of MIS to the distinctive requirements of the company (Grant, 2010).

4 RECOMMENDATIONS

From the discussion above, it is obvious that there are so many benefits associated with proper MIS (Heijden, 2009). The only question that remains is how any organization can ensure that it has proper MIS in order to ensure it benefits fully. The first thing to look for is the kind of leadership the organization has. Any organization must have a leadership that sets a perfect example that shows the importance of proper MIS. The management must therefore carefully select leaders that have the interest of the organization in question at heart and therefore form a team that leads by example to show how effective MIS can be beneficial to the organization economically, among stakeholders, security-wise, among workers and so on.

Similarly, the organization must ensure it has a proper plan on how to implement a good MIS (Shajahan, 2007). For example, the organization must ensure that it has all the required resources for the implementation and sustenance of good MIS. The resources can include relevant man power and financial capital to invest in modern technological requirements (Heijden, 2009). Specifically, the organization in question must hire experts in human resources management, information technology, production, and, generally, MIS matrix.

Specifically, experts in information technology must be hired so as to provide manpower to train other employees. Here the management must however be careful not to incur losses in the name of acquiring a particular form of MIS. For example, the management must perform a prior assessment to determine the size of the resources required,

possible outcome, the basal requirements of the organization in question, the profiteering chance (Heijden, 2009) and so on. In other words, the kind of the plan a particular management intends to implement must be within reasonable limits in terms of input, and the possible output since the bottom line for any organization is implementing only the processes that bring along positive change (Shajahan, 2007).

Capacity building and recognition of achievements is another important effort (Shajahan, 2007). The management must invest in regular training of the workforce at all level to ensure that they are up to date and keep up with the looming and rapid changes in technology and competitor's strategies. The regular training can be at all levels. The training must always revolve around the concept of MISs and the benefits thereof. In almost all industries, teamwork has become a key pillar for development. The capacity of the prevailing leadership to make and sustain coherent teamwork does not come easy; the management must ensure that it motivates its workforce. For example, the management must reward as well as punish achievers and offenders respectively with no favoritism. Similarly, offering free training and so on at particular times of the year must be among the many strategies that the management of an organization must employ while forming a proper MIS. Training is particularly essential in that it can be used to prepare the workforce to deal with any changes that might occur. Similarly, training offers the best chance to convince the workforce the stakes thereof when proper MIS is sustained. For example, MIS sustenance ensures sustenance of the workforce and the very essential protection of their rights. In the meantime, they are as well trained on how to handle and keep organization information with care.

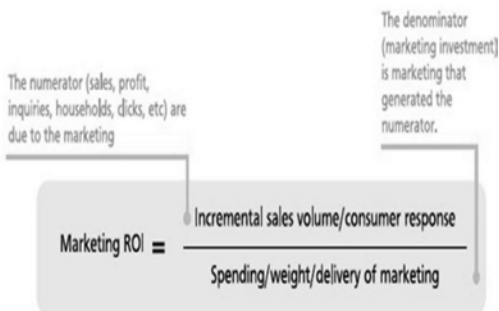
Doing research about what the competitors are doing is vitally important. This helps a great deal to keep in touch with the rate of technological changes. An organization must therefore make some study to determine its requirements and what the market has to offer at specific time period. It is not a must that such studies are done about the competitors only; rather, such studies can include other business or organizational entities in the market. Such studies help by providing a comparative basis where the organization in question easily locates the locus of its problems in the regard. Lastly, any organization must learn to keep vital information safe. Loss of information, leakage of information to unintended audience, and misplacement of information must always be avoided to avoid various related and/or legal problems. These are vital ingredients for a good MIS, necessary to keep the organization in question safe.

5 CONCLUSION

This study analysis has focused on the importance of MISs in organizations. It is evident that MIS affects organizations both internally and externally. The internal operations affected by MISs include communication, human resource systems and cost reduction, among others. On the other hand, the external organizational aspects affected by MISs include marketing and company image to name a few. These aspects underline the fact that MIS is more or less indispensable in the modern-day organization. It facilitates the identification and evaluation of the inherent relationships between organizations, people and technology. Decision-making has been highlighted as another pertinent area that is affected by MISs in an organization. This is because MIS facilitates the evaluation of information and data that pertains to all organizational aspects. The finding of the case study shows that the absence of an appropriate MIS at Sony Pictures Entertainment has undermined the company in numerous ways such as the recent hacking incident. As shown, assessing the organization situation and suggesting a proper MIS solution will positively affect the organization business performance.

6 FUTURE WORK

One of the most important tracking mechanisms for marketers relates to returns generated from investments. Tracking of sales and investment spending are (or should be) standard. One of early warning indicators is a Return on Investment (ROI). The ROI of Marketing activities (ROMI) is quite undefined to be measured against the strict objectives measuring business health, and thus ROI and ROMI may impact critical decision making to smaller degrees than hard and more concrete measures. We propose a formula for its calculation.



From the above formula the marketing ROI is calculated by dividing the attributable volume to marketing by investment of marketing. After applying and analyzing the marketing ROI on Sony Pictures Entertainment case study. It showed lowest marketing ROI due to reduce the sales volume and consumer response and using inappropriate MIS frame work. Comparing with the real estate business. It is one of the promising business opportunity with huge of marketing ROI in many countries, such as UAE, and the best investment with the highest return on investment in the long run.

REFERENCES

- Alec Crama W., Kathryn Brohmanb M., Yolande E. Chanb and Brent Gallupeb R. 2016. Information systems control alignment: Complementary and conflicting systems development controls. Volume 53 Issue 2 Pages 183–196.
- Ashlyn, S 2006. Outsourcing Management Information Systems. Hershey, PA: IGI Global.
- Bagad, VS 2009. Management Information Systems. Pune: Technical Publications.
- Clarke, S 2007. Information Systems Strategic Management. London: Routledge.
- Effy, O. 2008. Management Information Systems. Mason, OH: South-Western.
- Gibson, D. 2010. Managing Risk in Information Systems. Sudbury, MA: Jones & Bartlett Learning.
- Grant, K. 2010. Strategic Information Systems Management. OH: South-Western.
- Gupta, H. 2010. Management Information Systems. New Delhi: Hitesh Gupta.
- Haag, S. 2012. Management Information Systems for the Information Age. New York, NY: McGraw-Hill Higher Education.
- Heijden, M. 2009. Designing Management Information Systems. Oxford: Oxford University Press.
- Kayworth, T. 2008. Global Information Systems. London: Routledge.
- Kelkar, S. 2003. Management Information Systems: A Concise Study. New Delhi: PHI Learning Pvt. Ltd.
- Kennedy, G. 2005. Managing Information Systems: An Organizational Perspective. Financial Times Prentice Hall.
- Laudon, KC 2014. Management Information Systems: Managing the Digital Firm. Ontario: Pearson Education Canada.
- Sungjune P., Antonis S., Chandrasekar S, and Yuan N. 2015. Information technology and interorganizational learning: An investigation of knowledge exploration and exploitation processes Volume 52 Issue 8 Pages 998–1011.
- Shajahan, S. 2007. Management Information Systems. New Delhi: New Age International.
- Watson, H. 2015. Management Information Systems. Hoboken, NJ: Wiley.

Social benefits of innovative public administration services

Barbara Skoczyńska-Prokopowicz

University of Rzeszów, Rzeszów, Poland

ABSTRACT: The paper presents the results of the analysis regarding the application of e-New Public Management concept and assessment of the benefits of its implementation for the society and businesses in Poland. The article consists of an introduction, four parts, and a summary, which discuss the rationale behind the New Public Management concept, four models defining that management, as well as the consequences of implementing the NPM in the public service sector and in the area of social policy. The author defines the notion of governance as a consequence of implementation of the NPM, and points out another one, related to technological development, i.e. e-Public Administration Management, which enables providing public services electronically. In carrying out the principles adopted in the European Union, Poland created an organizational base, including legal regulations regarding informatization, to introduce digitalization and digitization to public administration. The content also includes definitions of e-service models and five levels of so-called maturity to render them. The next step includes building the Electronic Public Administration Communication System (SEKAP) and Public Administration Service Electronic Platform (ePUAP), improving interoperability, integrating SEKAP and ePUAP, as well as implementing the e-learning platform. All those actions make it easier to take care of official and administrative matters both for entrepreneurs and natural persons. The system requires further enhancement, however the benefits are already significant and noticeable. The author points out that Poland implements the EU guidelines regarding implementation of modern public administration management methods, it is in the process of intense digitalization, and the expected, real effects will be visible in a few years. Many projects will be carried out in the years to come, but already today the society can feel the changes which make it easier to take care of any matters in public offices, between offices, and in contacts at the meeting of business and office.

1 INTRODUCTION

Growth in innovation, technological progress, higher level of education in comparison to the previous years, continuous changes in the international arena, urbanization, as well as the development of the free-market economy, competitiveness and entrepreneurship contribute to continuous social and economic development of countries. The economic growth is accompanied by increasing needs of the society, requirements and expectations on the public administration, as well as on those in power and their actions. Societies need innovations, novel solutions and methods adjusted to their standard of living and expectations related to their states as institutions. Innovation, which became the “European culture”, must include not only the state and state organizations, but also entrepreneurs, citizens and administration.

The paper presents the results of the analysis regarding the application of the e-New Public Management concept and assessment of the benefits of its implementation for the society and businesses in Poland. The article consists of an introduction, four parts, and a summary, which discuss

the rationale behind the New Public Management concept, four models defining that management, as well as the consequences of implementing the NPM in the public service sector and in the area of social policy. The author defines the notion of governance as a consequence of implementation of the NPM, and points out another one, related to technological development, i.e. e-Public Administration Management, which enables providing public services electronically. In carrying out the principles adopted in the European Union, Poland created an organizational base, including legal regulations regarding informatization, to introduce digitalization and digitization to public administration. The content also includes definitions of e-service models and five levels of so-called maturity to render them. The next step includes building the Electronic Public Administration Communication System (SEKAP) and Electronic Public Administration Service Platform (ePUAP), improving interoperability, integrating SEKAP and ePUAP, as well as implementing the e-learning platform. All those actions make it easier to take care of official and administrative matters both for entrepreneurs and natural persons. The system

requires further enhancement, however the benefits are already significant and noticeable.

2 NEW PUBLIC MANAGEMENT

The directions of changes and reforms of the public administration have been discussed by scientists and practitioners since mid-1980s. Attention is paid especially to aspects related to exercising authority, control systems, and efficient functioning of the administration. Implementation of a traditional model of administration resulted in increased public expenditure, larger pressure exerted by citizens on solving social problems more efficiently, frustration, and a more critical approach to public authorities (Szumowski 2014, p. 92). The reasons for the origination of the New Public Management concept include mainly the inefficiency of Weber's model. The main allegation made against that theory is the strict hierarchic structure, formal requirements for applicants for officials, failure to take into account their practical skills, and many functional gaps. The new managerial approach to public administration appeared first in Great Britain, Australia, and New Zealand in 1980s and, in early 1990s, in the United States (Supernat 2003).

Studies do not include one consistent definition of the new approach to management. However, there are at least four approach models. Model 1 (Efficiency Driver) focuses all activities on efficiency. Efficiency in "New Public Management" is understood mainly as providing citizens with access to services of highest possible quality which are characterized by a relatively low cost for the state budget. Model 2 (Downsizing and Decentralization) calls for limiting the influence of the state and decentralization. Model 3 (In Search of Excellence) seeks perfection, excellence in public management. Finally, model 4 (Public Service Orientation) focuses on public services (Nawojczyk 2015, p. 172). New Public Management is specified by some authors as the modern form of the managerial approach to administration. Its main feature is implementation of public administration principles that govern the free market, especially setting a target consisting in ensuring improved efficiency of the functioning of administration by using mechanisms known from the private sector (Ochnio 2012, p. 4).

New public management means focusing all efforts on a citizen and customer referred to as an applicant, but also the public sector becoming similar to the business sector. The most important consequence of implementation of New Public Management was creating internal markets, based on competitiveness principles, for goods and public services. Rules formulated in this manner aim to draw more attention to customers-applicants,

offering a wider catalogue of goods and services, and a higher standard of benefits. All postulates promoted by new public management are implemented by: "streamlining" the government administration, delegating authority and responsibility to lower management ranks, and replacing ruling with co-management, directing the process of taking decisions to the mission and set goals, replacing distribution of government funds with actively seeking alternative sources of financing, creating the set of standards for implementation of public tasks, measuring effects of actions by the level of citizens' satisfaction, promoting the market mechanism at the expense of bureaucratic mechanisms (Osborne, Gaebler 1994, p. 511). It is worth mentioning that the essence of the new public management consists in concluding contracts for rendering services with private companies. There are many benefits of concluding transactions in this way, especially revealing costs of services and comparing cost effectiveness and efficiency. One must not forget about one more important benefit of concluding contracts, that is the lack of the monopolistic position of budgetary units and facilities, which contributes to limiting pressure groups and trade unions (Dunleavy 1986, p. 16). Implementation of the efficiency standards for the public administration entails the need to perform assessment of officials, who are rewarded for desired results within the concept. The principle was introduced of linking promotion to results, which entails increasing efficiency of work and prevents excessive employment in administration, provided that standards and norms of work are specified appropriately.

New Public Management obligates the state to seek an optimum relationship between the citizen and the state, and this relationship is evident in shaping and the area of the social policy. Non-profit social organizations, local communities and neighborly groups took over a large part of tasks within providing services for people in need, the handicapped, the elderly, families affected by dysfunctions and poverty. Involvement of citizens in such initiatives aimed to transform passive beneficiaries of goods and services into active participants of decision-making processes and participation in initiatives in which they had not participated before. Positions of involvement an active co-operation entail a number of benefits, especially expending resources more rationally and reinforcing the democratization principles (Lehning 1998, p. 220ff).

The literature often contains the notion of governance. It is a notion with a wide meaning, referring to the role and possibilities of the state or public authorities in the scope of shaping, enabling and taking actions to promote such social goals which are not sufficiently accomplished by the market and civil society (Nawojczyk 2015, p. 174). This wide definition specifies the range of governance. The

term co-management would be a suitable equivalent of that notion. It was noticed that the growth in citizens' expectations and the need for comprehensive services in the public sector result in dissatisfaction and the need for changes, therefore the alternative is co-participation of citizens in some forms of ruling and carrying out social needs. This is accomplished by coordinating collective actions at all levels from local, through metropolitan and regional to the national one. Co-management takes place at many layers and engages entities by encouraging co-operation in the form of partnership, negotiations and horizontal connections, increasing efficiency and ability to cope in the conditions of growing expectations. Public governance will refer mainly to civil society, defined as a network of social organizations. The role of authority defined in this way is limited to creating conditions, or frameworks, managing networks, solving related problems. In this concept, frameworks for the development of democracy are created, and democracy is not just limited to elections, but instead it is about being active, making decisions together, and creating a shared common good. Public governance activates citizens and encourages stockholders' involvement, constituting participating democracy. According to that concept, public tasks which can be carried out are entrusted with social entities, while retaining responsibility for providing public services (Kowalczyk 2008, p. 9).

3 E-PUBLIC ADMINISTRATION MANAGEMENT

The society and the state keep developing, and this entails the need for new solutions, implementation of new technologies, and the need for changes. Owing to the digital revolution and changes in the society, the electronic administration services are beginning to play an increasingly important role, both for the state and the citizens. Public Administration Management is not new, as the need for public administration management was born already in the previous century, however E-Public Administration Management is a novel concept. Plans of the European Union within providing public services electronically were specified in detail in the European Plan of actions for e-government for the years 2011–2015. E.g., according to the plan, 50% of citizens of the EU and 80% of enterprises will be using public e-services by 2015. The plans are part of actions supporting the European Union in carrying out the strategy Europe 2020 (14.COM (2010) final 2020). Moreover, a new word, 'digitization', appeared in Polish and in legalese. It was popularized when the new Ministry of Administration and Digitalization was established in Poland in 2011 to deal with a wide range of issues,

from religious minorities to informatization. When the notions of digitalization and digitization were introduced, doubts began to emerge regarding definitional vagueness, as some people began to use those words interchangeably. However, the word digitalization can be used in different meanings (17).

The notion of E-Public Administration Management is related to the notion of digitalization, but also digitization. Digitization means mainly changing the form, transforming a real, physical world, one or many of its components into a digital "equivalent". From the point of view of economy, digitization means better ergonomics, that is easy, quick access, propagation, as well as saving "space" (18).

Digitalization is very difficult to define—the word is used in many contexts rather liberally. The definition used in the act, i.e. the one used by the Ministry of Digitalization is the one that is the most credible. In the program, "Digital Poland", the ministry defines three fields of digitalization: access to fast Internet, development of e-services available on the internet, development of e-services and resources available on the Internet and competences of citizens' digital skills. Under the entry "Key actions of integrated informatization", it lists tasks such as ensuring organizational frameworks of the construction of the state information system, obtaining interoperability of public registers, or starting the State Calculation Cloud. It is therefore evident that the notion of digitalization is defined widely as actions in order to increase access to the Internet and its resources for citizens and building electronic mechanisms into the state administration. Consequently, there is a clear definitional discrepancy in the two notions of digitalization and digitization, which are commonly used interchangeably (17).

The use of the new information processing technologies is a vital element of the functioning of the public administration in Poland. For administration acting based on the constitutional principle of law and order, there is a need to use appropriate legal bases for those technologies. Poland is still implementing the "electronic administration". At the moment, the Act on Informatization of Activities of Entities Performing Public Tasks is applicable (Act of 17 February 2005). Implementation of legal solutions within this scope aims to bring public administration into 21st century. Informatization of the state means, in fact, informatization of the administrative function of public entities acting on behalf of the state and represents supplementation of other legal regulations from the scope of public orders, access to public information, and even civil service (Wiewiórowski 2010).

Appearance of legal regulations regarding informatization became a basis to use new ICTs

(Information and Communication Technologies) in public administration. All actions aimed at informatization will be referred to as e-administration, or e-office. For lawyers, e-office, or e-government, refers not only to public administration, but also to other entities of public authority. According to many definitions, e-government uses ICTs to improve the quality of the functioning of offices (19). It is essential to provide necessary services related directly to the ICT sector and satisfy the following needs: social—satisfied by specific institutions (e-health, e-safety, e-education), economic—satisfied thanks to business activity (e-services, e-work, e-commerce), informational—related to getting knowledge and social, official communication (e-administration, e-economy) (Kasprzyk 2011, p. 343).

The development of e-government means providing access to e-services publically. The services are a result of the existence of different relations in public administration institutions and at the meeting of administration with the surroundings. External relations include co-operation of public administration offices and entrepreneurs (G2B, B2G) and relations between public administration offices and citizens (G2C, C2G). On the other hand, there are internal relations between public administration offices (G2G) and offices and their employees (G2E, E2G) (Ziemia, Papaj, Będkowski 2013, p. 427–446).

Since e-government is a very extensive subject, there is no uniform division of the notion, and the classification will depend on the criterion used. Most often, the following e-service models are listed: B2B—Business to business, which is based on a business providing a service for other businesses, B2C—business to customer, when a business provides a service for individual customers, C2C—customer to customer, when an individual customer provides a service for other individual customers, C2B—customer to business, when a customer provides a service for businesses (Flis 2009, p. 12).

From the point of view of the subject, a division of e-services in the area of internet communication will be important, e.g.: a service enabling communication by mobile phones by means of the bluetooth protocol—a free-of-charge alternative of communication at a very close distance—C2C model, then video-streaming services, defined as videoconference, teleconference or on-line transmission services and virtual meeting rooms, where documents can be exchanged—B2B model; a search engine making it possible to look for people on the Internet, enabling to avoid information overload by filtering out redundant data—B2C model; advertising services on the Internet, a classic form of matching buyers and sellers of tangible and intangible goods—mixed B2B/B2C/C2C/C2B model.

According to many studies by professors dealing with the subject-matter of administration,

e-services can be provided at five levels of maturity. The first and basic level of maturity regards information: it means that public administration institutions give access for citizens and entrepreneurs to public information on Internet portals. Another level is the interaction level which consists in communication of stakeholders with offices electronically, but only one way, whereas the next level will mean two-way relations. The fourth way of maturity is referred to as transactional, because it is related to the possibility of carrying out all actions necessary to take care of official matters electronically. The fifth level—customization, enables dealing with an official matter electronically and at the same time introduces customization of service (Flis 2009, p. 12 ff).

4 GOVERNMENT TO BUSINESS AND IN E-PUBLIC ADMINISTRATION MANAGEMENT

The fact that we are an IT society, as well as changes and development of technologies mean that businesses and enterprises are also subject to changes and evolution. According to the conducted examinations, the most developing area of business, where e-services play a key role, is servicing companies, including professional business services, financial agency, electronic banking services, cloud computing and web solutions, electronic commerce, training and increasing workers' qualifications, referred to as e-learning. It is noteworthy that those types of activity are related mainly to the necessity of reducing costs and saving time. In the times when comfort and time are at stake, businesses care about carrying out transactions, accounting, tax settlements, investment consulting and other services, quickly and without having to leave their premises.

The Internet is becoming another sales channel for commercial entrepreneurs. They do offer their products with traditional methods, but increasingly also through Internet shops and electronic auctions. Over the last 10 years, in response to the demand for e-services in business, equivalent to relevant e-service models, many portals were set up to assist small, medium and large enterprises in implementing new technologies. For example, we can list the portal *biznes-firma.pl*, which supports running a business by means of the internet in e-commerce, e-services, which is a response to the growing demand for consulting services—B2B model. Another portal is *ibuk.pl*, which deals with sales of audio and e-books and magazines in the electronic version together with an internet publication—B2C model. The portal *bookkeeper.pl* runs internet accounting services for people running their own business. It is also worth mentioning services offering prices comparisons

between shops, such as ceneo.pl or skapiec.pl, and services presenting financial information, offering various financial calculators, e.g. bankier.pl (Batko, Billewicz 2013, p. 50).

Few people realize that efficient public administration is the key factor in the development of e-services. That is why increasing the role of management as well as building and developing public administration are so important. In this scope, the following can be distinguished: government within public administration offices—A2 A, co-operation of public administration offices and enterprises A2B, B2A, between public administration offices and citizens A2C, C2A (Batko, Billewicz 2013, p. 50ff).

Designing and implementing ICT systems supporting services provided by public administration units is a complicated issue, especially due to the need for organizational and legal changes and defining and implementing standards including units participating in that project (Batko, Billewicz 2013, p. 57). It is worth paying attention to the practical aspect of e-government in public administration, that is the SEKAP and ePUAP systems.

The Silesian Voivodship was the first one to implement the program constituting an exemplification of e-government in Poland—the Electronic Public Administration Communication System (SEKAP). The program was to be carried out in the years 2005–2008 and enabled providing public e-services at various levels of maturity in relations G2G, C2G/G2C, B2G/G2B. The next step was to introduce the project Development and Popularization of the Electronic Public Administration Communication System in the Silesian Voivodship—SEKAP2. The new program enhanced the previous one by increasing the number and quality of public e-services, enhancing interoperability, integration of SEKAP and ePUAP, as well as implementation of the e-learning platform. The basic goal was to organize the cycle of training sessions both for residents, and public administration officials (Ziemia, Papaj, Bedkowski 2013, p. 427–446).

Another system which is essential for increasing the quality of electronic services in public administration is the Electronic Public Administration Service Platform (ePUAP). The ePUAP is an all-Polish ICT platform, by means of which public administration units and public institutions can provide their services electronically, i.e. creating and servicing electronic documents, sending electronic documents, exchanging data between ePUAP and other ICT systems, identifying users and ability to settle their activities, verification of the electronic signature, creating public entity services, servicing electronic payments, conforming the trusted ePUAP profile. It is an IT system, thanks to which citizens can take care of official matters

on the Internet, and representatives of public entities can make their services available free-of-charge electronically. The idea behind building ePUAP was to create one, available and safe place to provide electronic public services, giving access to public e-services type C2G/G2C, B2G/G2B and G2G. The functioning of the system goes back to the years 2006–2008 within the project Building Electronic Public Administration Services Platform—ePUAP. In the years 2009–2013, the program ePUAP2 was carried out to develop the functionality of the portal by increasing the range of services provided electronically (Ziemia, Papaj, Bedkowski 2013, p. 427–446).

The operation of those two systems for entrepreneurs has great consequences. First of all, it facilitates taking care of all matters related to establishing economic activity. It is worth remembering that informatization of administrative structures developed to such an extent that a limited liability company's registration documents can be submitted electronically by means of an internet portal and the registration takes place within 24 hours of submitting the application. Second of all, ePUAP can be used to submit all applications and take care of official matters by means of completing electronic forms. The advantage of that system is its speed, efficiency and convenience for users. Over the years, both systems have developed a lot: the range of rendered e-services has been extended and enhancements have been introduced for users' convenience.

5 BENEFITS FOR ENTREPRENEURS

There are countless benefits of introducing all kinds of e-services. At the moment, the Internet is one of the main work tools and information access tools, so it should also be used to the full to take care of official and administrative matters. In implementing such solutions, state authorities were guided by universal motives such: saving time, wide scope of possibilities and varied offer of services, comfort, breaking geographical and time barriers, financial savings. Electronic systems were introduced with natural persons in mind, to enable them to electronically take care of common administrative matters, such as: changing permanent address, passport-related matters, IDs, accessing official information including the Register Office, submitting tax declaration, using work agency services, registering the unemployed and job-seekers. Advantages of the system include its consistency and multitude of offices using electronic systems, as well as increased safety by limiting the possibility of entering the same data many times. For entrepreneurs, implementation of electronic solutions is important mainly for the

following reasons: availability of all necessary data to establish or run activity in one place, possibility to take care of and check the case status at any time and place, saving time and speed in taking care of a matter, unlimited office hours, saving materials for companies, order and using one database of documents necessary to use public administration services (Kasprzyk 201, p. 344).

5 CONCLUSIONS

Over the recent years, the model of managing public matters in Poland has changed. Many new solutions and initiatives were undertaken to activate the society, by carrying out the postulate of shared participation in decisions regarding matters which directly affect citizens. Informatization of offices and courts significantly contributes to those changes: electronic protocols were introduced, European style electronic writ proceedings, documenting the course of court proceedings by means of ICT. Poland is changing, at an accelerated pace becoming similar to developed countries of the Western Europe, among other things thanks to the use of EU funds, by developing increasingly effective public administration management solutions, and activating individual citizens and the whole society. Changes introduced in the public administration create a new legal situation also for entrepreneurs, who are able to have their needs satisfied faster and with the use of means which were not used previously. The constant development affects enterprises by enabling them to take care of their matters faster, more smoothly and conveniently. The above analysis indicates that Poland implements EU guidelines regarding implementation of modern public administration methods and it is at the stage of intensive digitalization, whereas the expected, real effects will be visible in a few years. Many projects will be carried out in the years to come, but already today the society is affected by the changes, which facilitate taking care of any matters in public offices, between offices and in contacts at the meeting of businesses and offices.

REFERENCES

Act of 17 February 2005 on Informatization of Activities of Entities Performing Public Tasks, *Journal of Laws* Dz.U. 2005 no. 64 item 565.

Batko K., Billewicz G., E-usługi w biznesie i administracji publicznej, *Studia Ekonomiczne, Uniwersytet Ekonomiczny w Katowicach* 2013, [online], http://www.ue.katowice.pl/fileadmin/_migrated/content_uploads/3_K.Batko_G.Billewicz_E-uslugi_w_biznesie.pdf. [Access on 29.02.2016].

COM (2010) 2020 final, EUROPE 2020 A Strategy for Smart, Sustainable and Inclusive Growth, European Commission [online].

Dunleavy P., Explaining the Privatization Boom, "Public Administration", vol. 61/1986.

Flis R., E-usługi-definicja i przykłady. Badanie zapotrzebowania na działania wspierające rozwój usług świadczonych elektronicznie (e-usługi) przez przedsiębiorstwa mikro i małe, www.parp.pl [online], http://www.web.gov.pl/g2/big/2009_12/e128419bc4aca1881822862d9da143f5.pdf.

http://europa.eu/legislation_summaries/information_society/strategies/index_en.htm, [Access on 29.02.2016.]

<http://www.pilsudski.org/portal/pl/466-digitalizacja-acyfryzacja>, [Access on 29.02.2016].

<http://networkeddigital.com/2014/05/17/definicja-cyfryzacji/>, [Access on 29.02.2016].

<http://www.edukacjaprawnicza.pl/artykuly/artikul/a/pokaz/c/artikul/art/informatyzacja-administracji-publicznej-w-polsce.html>, [Access on 29.02.2016].

Kasprzyk B., Aspekty funkcjonowania e-administracji dla jakości życia obywateli (in: Nierówności Społeczne a Wzrost Gospodarczy), Zakład Metod Ilościowych, Wydział Ekonomii Uniwersytet Rzeszowski 2011, [online], file:///C:/Users/domka_000/Downloads/027.pdf, [Access on 29.02.2016].

Kowalczyk L., Współczesne zarządzanie publiczne jako wynik procesu zmian w podejściu do administracji publicznej/*Zeszyty Naukowe Wałbrzyskiej Wyższej Szkoły Zarządzania i Przedsiębiorczości*, nr 1/2008/, ISSN 2084-2686.

Lehning P. B., Towards Multicultural Civil Society: The Role of Social Capital and Democratic Citizenship, "Government and Opposition", vol. 33/1998, p. 221-242. ISSN 0017-257X.

Nawojczyk M., Nowoczesne formy zarządzania w administracji publicznej, *ZN WSH Zarządzanie* 2015(2), p. 169-181, ISSN 1899-8658.

Ochnio M., Nowe Zarządzanie Publiczne (New Public Management)—podstawowe cechy modelu. Jego zastosowanie w Polsce, *Stowarzyszenie Instytut Zmian*, Warszawa 2012, [online], http://iz.org.pl/wp-content/uploads/2012/09/New_Public_Management_Michal_Ochnio_Institut_Zmian.pdf.

Osborne D., Gaebler T., *Rządź inaczej. Jak duch przedsiębiorczości przenika i przekształca administrację publiczną*, Media Rodzina, Poznań 1994. ISBN 838594132.

Supernat J., Administracja publiczna w świetle koncepcji New Public Management, *Zakład Nauki Administracji, Uniwersytet Wrocławski*, [online], http://www.supernat.pl/artykuly/administracja_publiczna_w_swietle_koncepcji_new_public_management.html, [Access on 19.02.2016 r].

Szumowski W., Zarządzanie publiczne- próba systematyzacji koncepcji, *Uniwersytet Ekonomiczny we Wrocławiu, Nauki o zarządzaniu management sciences* 4(21) 2014, p. 86-98, ISSN 2080-6000.

Wiewiórowski W.R., *Zasady leżące u podstaw administracji publicznej*, *Pracownia Informatyki Prawniczej Uniwersytet Gdański* 2010, [online], http://arch.prawo.ug.edu.pl/pdf/zaklad3/Zasady_lezace_u_podstaw_informatyzacji_administracji_publicznej.pdf, [Access on 29.02.2016].

Ziemia E., Papaj T., Będkowski J., Egzemplifikacja e-government w Polsce-analiza porównawcza SEKAP i ePUAP, *Roczniki Kolegium Analiz Ekonomicznych*, nr 29/2013, p. 427-446, ISSN: 1232-4671.

The use of Information and Computer Technology in the decision-making process of foreign students concerning their future professional career

Barbara Sobiczewska

Faculty of Social Sciences, Institute of Psychology, The John Paul II Catholic University of Lublin, Lublin, Poland

ABSTRACT: More and more dynamic development of state-of-the-art information technologies carries with it a number of changes. One of many fields in which the IT software solutions may be applied includes, among other things, personnel management. Companies make use of the latest solutions of the IT branch in the process of searching for the potential job candidates and their recruitment. On the other hand, candidates may monitor the possibilities of their employment by looking for job offers online and if found some of them attractive may directly send application documents and references to the employer. Some companies implement their own employment platforms, while others make use of already established employment-related sites. The labour market is not territorially limited what gives the possibility to find future employees on an international level and choose people who seem to be most qualified and experienced in a particular field. Transnationalization of labour markets determines a tendency to plan and implement professional careers and educational migrations in an international and intercultural context. Job search websites, databases, and social media is a starting point for young people who just begin their employment career. ICT eliminates barriers and gives a possibility to enter job market anywhere in the world.

The purpose of this article is to show the degree of the use of Information and Computer Technology in the decision-making process of foreign students concerning their future professional career. The article shows a degree to which students are aware of the possibilities of modern Information and Communication technologies in the process of searching for job as well as in their professional development.

1 DEVELOPMENT OF ICT INDUSTRY

The societies of today's world base their development on information and communication technologies and transform themselves into information societies (Kurzyjanski, Pawelczyk, 2011). It is indicated that the information and communication technologies facilitate the process of reshaping the societies from industrial to information which base all aspects of social organisation on exchange, use and integration of information and knowledge reducing in this way the distance between nations (Kisielnicki, 2014). Modern IT technologies drive changes in many ways. There is, in particular, no discipline in which development would be made without the use of IT solutions. Easier flow of information, reduced costs of communication and transport lead to the increase in competitiveness of companies—companies that operate locally feel free to function on a global market. The organisation of work also changes, through e.g the use of new business models and increase in the number of

digitally available products. Ways of communications also transformed. Companies, employers and people who make use of the high-tech tools make use of the two-way communication which is easier and more efficient. Employers and candidates exchange information also by using traditional media. (Zajac, 2012).

2 ICT AND CAREER MANAGEMENT

One of the fields in which ICT is used is among other things the personnel management. The companies use the latest information technologies in the context of recruitment and selection, training of employees, and internal communication in enterprises. They constantly increase their innovation, and creativity of their actions. Employers through the use of ICT tools in recruitment and selection of employees have the possibility to reduce costs related to publication of advertisements in other mass media and accelerate their

search. Labour market is no longer restricted to a particular region. ICT tools give the possibility to employ people on an international level and find those who are best. The process of selection is more objective and may cover a bigger number of candidates. (Zajac, 2012).

On the other hand, potential candidates have an easier access to job offers and particular companies profiles as they are published on local job search websites and international portals as well. Furthermore, many job search engines offer searching by category, location, or even starting salary. Candidates may establish a tailored profile, upload a resume, and search job postings aggregated from company pages, associations, and various listings from across the Web. Through a direct access to job offers of particular companies it is possible to monitor what qualifications and features of employees are desired by employers (Zajac, 2012).

Employment portals destined for searching for jobs are most popular recruitment method in Europe and United States. There are a few categories of employment-related websites. Some of them publish job offers of various companies e.g. www.pracuj.pl, www.infopraca.pl. (in Poland), other have been developed both for candidates and an employer to exchange information www.indeed.com. other act like a search engine in the context of job search: www.monster.com. Specialised services sometimes focus on specific trades, e.g. only on offers in sales, manufacture or IT. Other publish advertisements concerning part-time or contract employment. Very often companies establish their own accounts at social networking websites such as LinkedIn, Goldenline, or Facebook and publish there their job offers. It must be emphasized that a social site as mentioned above is a good source of information about a particular candidate as an employer may view a personal profile of a person, see his or her interests, pictures and personal information, comments of other people, achievements etc. Sometimes companies employ people who search through the resources of a social networking website to find appropriate people. Nowadays, the way of searching for a person may be seen as informal, however, it is useful and more and more companies implement it as one of the stages of the process of recruitment and selection. Social networking websites are very popular among young and active people who use them as a form of communication and a life diary. Apart from social media, employer use their own communication channels, e.g. classifieds websites, business blogs, company profiled websites and many more (Zajac, 2012).

Modern IT technologies also facilitate the process of recruitment and selection as now it is possible to conduct teleconferences with candidates, see

application documents and carry out tests online (Zajac, 2012).

Undoubtedly related to all social media, including Facebook, LinkedIn, Profeto or Goldenline is the idea of networking as it is a process of making connections with other people and exchange ideas. It is based on creating networks of communities and cooperatives engaged in sharing experiences, resources, in learning and education, and in personal growth. A person through getting new contacts, broadening the scope of people that share similar interests and knowledge, and by sharing information creates a detailed and perfect professional image. Web platforms also make possible to have an access to information about conferences, meetings, lectures, events related to a specific business and participating in discussions in groups or exchange information with other users of network. (Jobvite: Social Job Seeker Survey, 2011).

3 SITUATION OF STUDENTS STUDYING ABROAD

Various researchers discuss the causes, progress, results of the contemporary problem of migration. (Górny and Kaczmarczyk, 2003). Some, analyse it from the psychological point of view. One of the studies focuses on the factors, determinants of being ready to leave home country to start studies abroad or find job. Here, the key focus is on preconditions of the decision to leave. As an exemplary study in this respect one may enumerate the study of Goździewicz (Goździewicz, 2013) carried out based on conceptual model of determinants and results of the openness to international carrier. Other, A. Bańki (Bańka, 2005) aims at specifying personal and contextual conditions of the final year students to start an international career.

Higher studies are treated as a basis for investing in own social and intellectual capital. An additional form of investment is participation in international education (Kawczyńska-Butrym, 2014). Transnationalisation of labour markets determines the tendencies to plan and realise professional careers and educational migrations in international and intercultural context. It constitutes a starting point of young people in other countries and societies and a possibility to stay (temporarily or permanently) abroad (Bańka, 2006).

International mobility of students and higher studies graduates gives a chance to continue getting knowledge or job in an international environment, learn new languages and culture. Bolognese experts claim that participation in international exchange of students and undertaking employment

abroad allows to understand a particular discipline of science in other educational system than domestic.

It also increases the international dimension of knowledge and its application, it allows to improve language competences and foster the development of skills necessary for working in multicultural environment often characterised by totally different social conditions (Zygierewicz, 2014).

In connection with the development of information and computer technologies it seems that the access to information about labour or studying opportunities abroad is still unexploited. In view of a diversity of offers, their competitiveness, and comparability anyone should have at least an access to platform presenting and comparing various possibilities though one website or application combining job search functions or studying search. Single websites, portals and social platforms are just a single point of knowledge for a particular person. Still, there are students who are studying in a foreign country, however, the extend to which they are aware of the opportunities the information and computer technologies offer in this matter was the subject of research.

4 THE RESEARCH METHODOLOGY

As it was previously mentioned, the subject of the research was the use of Information and Computer Technology in the decision-making process of foreign students concerning their future professional career. The research aimed at determining the degree of awareness of the possibilities the latest ICT give in the process of looking for job abroad or in determining the path of professional development of students.

The research was carried out in February 2016. The research group consisted of 30 first and second year students from Ukraine who were studying at one of the universities in Lublin, in Poland. The faculties at which they were studying included journalism, social communication and international relations. The group consisted of 19 women (63,3%) and 11 men (36,7%), of the average age—18, 37.

As a research method the diagnostic survey was applied with a survey questionnaire as a research technique. The questionnaire consisted of 21 questions. The respondents had to choose one out of a few suggested answers (a, b, c, d, e), mostly consistent with their beliefs. The questions concerned among other things, the frequency of using web portals and ICT tools, the purpose of using ICT tools, or opinion of students about usefulness of modern information technologies in the process of labour search.

5 RESULTS OF THE STUDIES

First of all it was verified on what source of information the students made decisions concerning starting higher education abroad. The results indicated that among various sources of information (such as foreign universities websites, educational websites, social networking sites such as Facebook, LinkedIn, internet forums, lecturer or counsellor in a domestic country or recruitment agency) the most often chosen information source were foreign universities websites (36,7% of the research participants). More than half of the students (60%) declared that they were searching the net in order to prepare for undertaking studies abroad, including finding information about the culture of a particular country, requirements and costs of studying there. More than half of the students applied for universities abroad by using (73,3%) the particular university website (establishing account, sending application and downloading required documents). A small percent of students (6,7%) made use of traditional sources (without using computer or Internet at all). In choosing university abroad the content of a website was the priority and main source of information for more than a half of students (60%).

As far as future professional careers of the research participants are concerned, it was investigated what sources of information the students would use in this respect. It turned out that the vast majority of the examined students (83,3%) declared that the basis source of information about labour market and employment offers is the Internet. Significantly smaller percent of people would use for that purpose the content of social networking websites (10%) (Table 1).

More than half of the students (63,3%) in thinking about a potential employer would search information about a specific company on this company's website. Considerably less students would search for such information in social media services

Table 1. Main sources of information, while looking for a job, according to students.

Answer	Frequency	Percent
a. Websites	25	83,3
b. Social networking services (Facebook, GoldenLine, LinkedIn etc)	3	10,0
c. Newspapers	2	6,7
Total	30	100,0

Table 2. Social networking services.

While looking for a job, I would also use:		
Answer	Frequency	Percent
a) LinkedIn	2	6,7
b) Goldenline	1	3,3
c) Euraxess	13	43,3
d) I would not use social networking services	14	46,7
Total	30	100,0

(20%) or internet forums (10%). Most participants (46,7%) would not use social media services during the process of looking for a job.

The research also included assessing a degree of knowing by students various social networking services and their usability. It turned out that more than a half of foreign students (76,7%) does not realize about the existence of LinkedIn, Goldenline or Euraxess (Table 2).

In using Facebook, the majority of the participants of the research (40%) belongs to a different community than related to a profession. A small number of students (10%) admits not having a profile on Facebook. Interesting were also results of the research concerning the degree of usefulness of IT tools (tablets, smartphones, internet portals, mobile applications) in searching for the opportunity to study abroad. The students generally assessed this usefulness as high (40%) or rather high (46,7). Rather low (10%) and low (3,3%) results are shown in Table 3.

The degree of usefulness of IT tools (tablets, smartphones, internet portals, mobile applications) in searching for the opportunity to develop professional career was assessed by majority of students as rather high (53,3%) or high (30%) (Table 4).

The majority of people evaluates its proficiency in using social networking websites as rather high (43,3%) or high (40%). The majority of the participants of the research use (43,3%) Internet portals, social forums and mobile applications as a source of knowledge on issues related to the field of study or as a way of getting in touch with family or friends (33,3%). There is also a small number of students who (13,3%) use them as a source of knowledge on the possibilities of the development of professional career or about the labour market. Some students find social networking websites as (43,3%) moderately helpful as far as finding job is concerned. Remaining students assess them as very (33,3%) or barely helpful (23,3%). Moreover, many foreign students believe that (60%) the employers often use social networking websites in their recruitment processes (Table 5).

Table 3. The degree of usefulness of IT tools (tablets, smartphones, internet portals, mobile applications) in searching for the opportunity to study abroad.

The degree of usefulness of IT tools (tablets, smartphones, internet portals, mobile applications) in searching for the opportunity to study abroad I assess as:		
	Frequency	Percent
a) high	12	40,0
b) rather high	14	46,7
c) rather low	3	10,0
d) low	1	3,3
Total	30	100,0

Table 4. The degree of usefulness of IT tools (tablets, smartphones, internet portals, mobile applications) in searching for the opportunity to develop professional career.

The degree of usefulness of IT tools (tablets, smartphones, internet portals, mobile applications) in searching for the opportunity to develop professional career, I assess as:		
	Frequency	Percent
a) high	9	30,0
b) rather high	16	53,3
c) rather low	4	13,3
d) low	1	3,3
Total	30	100,0

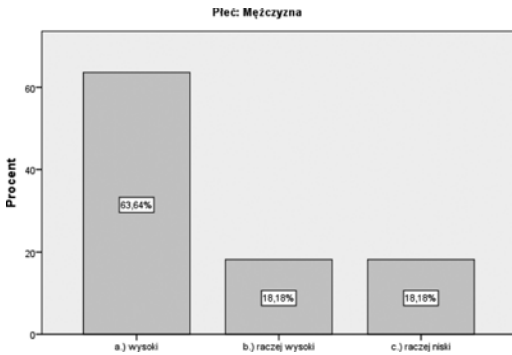
Table 5. Usage of social networking services for recruitment purposes.

I think that employers use social networking services for recruitment purposes:		
Answer	Frequency	Percent
a) always	2	6,7
b) often	18	60,0
c) sometimes	8	26,7
d) rarely	2	6,7
Total	30	100,0

The results of the studies reveal considerable differences between women and men in the assessment of usability and usefulness of computers, Internet and mobile applications in the process of searching information about the possibility of the professional development. The majority of women considers the rather useful (68,42%) or useful (26,32%)

On the other hand, the majority of men (63,64%) assesses the usability and usefulness of computers, Internet and mobile applications in the process

Table 6. The assessment of usability and usefulness of computers, Internet and mobile applications in the process of searching information about the possibility of the professional development made by men.



of searching information about the possibility of the professional development as high (63,64%) or rather high (18,18%). Little students consider it as rather low (Table 6).

6 DISCUSSION OF RESULTS

Based on the results of the study it definitely can be concluded that the decisions concerning professional or educational mobility of students was made by the students, who now are studying in Poland, on the basis mainly of websites of particular foreign universities and generally the content of other websites. There is really rather very small percentage of students who take into account other sources of information than the Internet and not need computers for their search. While thinking about future professional careers students also do not need anything more than the Internet and the content of sites related to labour market. Despite the fact that generally students believe that potential employers do have accounts and view the content of social media websites during the recruitment processes and searching for job candidates, the majority of foreign students (76,7%) do not know that services such as LinkedIn, Goldenline or Euraxess exist. The most commonly used social media website is Facebook that, in the majority of cases, is used not for professional but private information search. Generally, all IT tools (tablets, smartphones, web portals, mobile applications) are considered helpful in searching for the possibility to study abroad or during the process of searching for job abroad.

To sum up, the use of Information and Computer Technology in the decision-making process of foreign students concerning their future professional career is essential and constitutes the basis of future actions of the majority of students.

7 CONCLUSIONS

Students do realize that modern computer and information tools are extremely useful in today's world. Moreover, young people are conscious and use IT solutions in the decision-making process. It was owing to the content of university websites or generally information available online that they have undertaken a decision to study abroad. Generally, the Internet is a place where the search takes place. It is the main source of information. People make decisions concerning their future based on the content found in the net. It is a trusted and a reliable source of contemporary information. Social media services in the respect of finding a job or a place for studying are considered as "helpful" but still the possibilities many of such platforms is unexploited and not known to students.

Generally the participant of the research were foreign students who stay in a different country than in which they were born. The research group is definitely an exemplary group of young people who have made and will make decisions about their future employment as well as future educational and professional career. The participant of the research do use modern IT applications, search the content of many websites, use the social media services, and base their decisions on information available there but still their knowledge on the possibilities the ICT offers on account of labour market analysis and professional development should be extended.

Despite some limitations, the research may be an indicator for universities or employers in guiding and showing students how information and communication tools are useful in the process of professional development right from the beginning of the young people's educational and labour career. It is essential to extend students knowledge about such sciences as: social sciences, the psychology of professional development, intercultural career advice, and psychological theories of decision making. Lectures or classes with people who are able to show students how combine IT knowledge with efficient labour search or how to prepare a profile of a candidate for particular job in advance would be an accurate advice. Companies should also focus on creating a combined, intercultural and international platform enabling both the search for job offers worldwide, contacting employers and creating profiles for job candidates and popularizing it among students. If media services are popular and broadly used a similar professional social media service would be a fine solution. The knowledge obtained during this research makes an accurate counselling concerning global career, increasing competences, including intercultural ones.

REFERENCES

- Bańka, A. 2006. *Poradnictwo transnacyjne. Cele i metody międzykulturowego doradztwa karier*. Warszawa: Ministerstwo Pracy i Polityki Społecznej.
- Bańka, A. 2005. *Otwartość na nowe doświadczenia życiowe: podstawy teoretyczne oraz struktura czynnikowa Skali Otwartości na Karierę Międzynarodową*. Warszawa: Instytut Rozwoju Kariery.
- Goździewicz, A. 2013. *Otwartość na karierę międzynarodową studentów ostatnich lat studiów*. Gdańsk: Wydawnictwo Uniwersytetu Gdańskiego.
- Górny, A. & Kaczmarczyk, P. 2003. Uwarunkowania i mechanizmy migracji zarobkowych w świetle wybranych koncepcji teoretycznych. *Pracemigracyjne (49)*. Warszawa: CMR Working Papers. [online] <http://www.migracje.uw.edu.pl/publ/208/>
- Jobvite. 2011. *Social Job Seeker Survey 2011*. [online] <http://web.jobvite.com/rs/jobvite/images/Jobvite-Social-Job-Seeker-Survey-2011.pdf>
- Kawczyńska—Butrym, Z. 2014. *Migracje edukacyjne. Studenci zagraniczni—dwie strony księżycy*. Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej.
- Kisielnicki, J. 2014. *Zarządzanie i informatyka*. Warszawa: Wydawnictwo Placet.
- Kurzyjamski, R. & Pawelczyk, Z. 2011. Wykorzystanie Internetu przez studentów – wyniki badań. *Acta UniversitasLodziensis, Folia Oeconomica, (261)*. Łódź.
- Zajac, J. 2012. *Technologie informacyjne i komunikacyjne a zarządzanie personelem*. Publikacja przygotowana i wydana w ramach projektu badawczego pt. „Trendy rozwojowe i zmiany gospodarcze w regionie”. Warszawa: MGG Conferences. [online] <http://www.mgg.conferences.pl/media/pdf/reports/zarzadzanie-personelem.pdf>
- Zygierewicz A. 2014. Międzynarodowa mobilność edukacyjna studentów. *Analizy BAS 11 (115)*. Warszawa. [online] <http://bazekon.icm.edu.pl/bazekon/element/bwmeta1.element.ekon-element-000171278313>

Modelling of the process teaching-training in E-learning

A.T.A. Yalid

Laboratory of Physical Chemistry Analytic, Ben M'Sik Faculty of Sciences, Hassan II University of Casablanca, Casablanca, Morocco

M. Bassiri

Laboratory of Physical Chemistry of Materials, Ben M'Sik Faculty of Sciences, Hassan II University of Casablanca, Casablanca, Morocco

M. Moussted & M. Talbi

Laboratory of Physical Chemistry Analytic, Ben M'Sik Faculty of Sciences, Hassan II University of Casablanca, Casablanca, Morocco

ABSTRACT: The new approaches of educational and didactic screenplay of on line formation answer the problematic of individualization and differentiation of formation course.

The personalization of this formation course professionalizing requires the recourse to the process of modelling of the teaching proves to be like powerful vector of apprenticeship optimization at the adult learners, it is about modeler the activities of trainings aimed in objects of formation, that is sceneries the content of a formation module in numerous activities in order to be able to reconcile them with the requirements of the computerized context.

The modelling permits the generation of the different courses of training therefore.

The institutional instructions and the advisable educational practices founded on the approach by expertise that aims objectives of higher level [taxonomy of Bloom] (referring modelling).

It encourages the development of educational sequences resting the activities of analysis, assessment and synthesis.

The object of survey of this research of modelling of the educational scripts will be analyzed according to different disciplinary approaches (the sciences of the education, the didactics of the disciplines, sociology, the psychology of work, and the ergonomics, the sciences of the language, etc.). The teacher's activity consists in conceiving educational sequences adapted to the needs, to the styles of training of the learners adult ("converge, divergent, accommodator, assimilative") and to the requirements axiological (to enact the finalities).

In this perspective the teacher inventor is brought to categorize the different activities of training to create the favorable conditions of investment, engagement and mobilization of the expertise in a global and harmonious (cultural, technological, strategic, methodological and communicative) manner.

In this on line formation context, to conceive educational and didactic scripts come back to put an in place together ordered of activities governed by actors who use and produce resources of training "Paquette 07".

In our survey of case it is about numeric resources organized according to the aimed expertise and of the instrumental specificities of the supports of mediations.

Keywords: pedagogical scripts, pedagogical modelling; engineering of formation, referential of skills

1 INTRODUCTION

The reforms present the education and the formation attached a fundamental importance to the modelling of the methods and convenient educational in a quest of efficiency of the formation act. This modelling leans on the approaches educational "behaviorism, constructivism, cogni-

tivist socio constructivism" [Houssay 09] [1]. She/it constitutes a powerful lever of socio-professional qualification by the conception and the screenplay of the devices of formation and expertise.

Our object of research consists in proposing a modelling and conception of the educational scripts professionalizing integrating the numeric technologies. We define the objects of survey; the

educational activities correspond as the description of the organization and the progress of different units of trainings.

This descriptive survey and exploratory aims to articulate way very strong two measurements that are paradox: the theoretical dimension and the operative dimension. The crossing and the complementarity of these two aspects contributes to the setting in relief of a conceptual setting constructed of the models of reference, that encourage in a first time the exchanges between the theoretical foundations. “Epistemological” of the process of modelling of the objects of survey and that analyzes in a second time the steps pragmatic and operative recommended by different platforms that allow the users to appropriate these activities scenarists and to use them in an adaptive and efficient way.

The central question of research expresses itself of the following manner: facing the difficulties of location of the logics of construction and formalization of the clear and explicit educational activities. To what educational screenplay gait is she/it necessary to put in place in a course professionalizing?

2 THE REVIEW OF LITERATURE

The setting of reference that served to the analysis of our survey consists mainly of the knowledge descended of different fields. 1° the steps of engineering of formation and expertises. 2° des theoretical perspectives and phraseological of the professional didactics. 3° Modelisation of the systems of information “Modelling profession”, 4° l’ingénierie didactic and educational.

3 METHODOLOGY PRECONIZED

The general objective of this research work is to palliate to the problems of the classification, d hierarchization, of specification and diversification of the objects of studies and activities andrago-educational support of teaching—training of the models of the scénarizations educational. The descriptive gait and exploratory recommended has for goal to characterize the uses, the waiting and the needs of the teacher-inventors in terms of scenarisation ion. [Pernin & Emin 06; Villiot-Leclercq & Pernin 06] [2].

He/it permits the modelling of the processes of conception of the scripts of activity of the teacher-inventor in a gait of formation engineering professionalizing. Jointly this survey has for object the mutualisation and the capitalization of these educational scripts constructed. This methodological takes support on the contributions of the formation engineering [3].

V Psyché, O Mendes, 2003 and of expertise and more precisely of the engineering of the needs and the systems support of mediation. We also wish that our model appears in a perspective socio constructivist and interactive [philipe jenneart] [4].

He/it generated thus to a process of operationalization and descriptive formalization structuring the different scripts of training:

- To formalize models of conception and screenplay of the situations of formation professionalizing.
- The modelling of the mediatisation environments.

The present intervention titled “Modelling of the process teaching-training in E-learning” is an is a descriptive survey of the educational scenarisation that refers to the national educational norms (The approach and the referential of expertise as powerful lever of the reflexive teaching to the training), to which we add the methods of development in engineering didactics of the systems of information aiming to modeler the activities of on line training. This scenarisation allows the adult learner to structure and to organize their activities in a meaning manner. Beyond the modelling of the content scenarist, our object of survey also puts the focal on the classification of the activities of training and the different implied in this process in this perspective. The educational engineering and the techniques of modelling will permit the setting up of the graphic and symbolic representations of the models required to surround the complexity of formation from afar while conceptualizing an environment of development professionalizing.

We completed this methodological setting exploratory by an investigation of research by 12 teacher’s researchers intervening in the formation of the Masters ITEFS (engineering of the technologies for the education and the formation faculty of the sciences BEN M’SIK) and 26 academic students of this formation degree course. While especially debating the possibilities of reuse of the results observed in the formalization of the objects of studies and the educational activities. We interest mainly in the treatment of the data to the process of modelling and his/her/its integration in the educational scripts permitting rich and varied interactions between the educational community and the adult learners.

4 CONCLUSION

In this descriptive research work and exploratory, we are interested to conceptualization and the formalization of the educational scripts according to a referential profession oriented by the integration of the NTIC.

Table 1. Result of the questionnaire of the scenarisation of the

Questionnement in the studying Master ITEFS	Answers
Q1. The modes of training recommended are they adapted to the x styles of training of the adult learners	75% des reponses mentionées révèlent un décalage entre les procédés proposés et les démarches cognitives préférentiel des étudiants of training of the adult learners.
Q2. The computer instrumentation is it interactive to the solicitation and to the investment of the students.	65%: some respondents think that the design of the flat frame is sophisticated but moins interactive screw opinion of the waiting and the individual needs of the different learners (his/her/its caractère standardisé and normalized).
Q3. The educational mediation encourages t—it the solicitation collaborative of intelligent guidance. (Importance of work collective interapprenants)	88%: some students reveal the rigorous difficulties of accompagnement soliciting a good interactivity.
Q5. The scenarisation of the contents teaching is it adapted to the courses staff of the learners	78% of the learners think that the majority of the happy is not specific to the different runs of formation staffs of the students.
Q6. The Knowledge and expertises aimed by the scripts are them in consistency with the referential of the academic expertises (vertical consistency)	89% of the respondents explain that the aimed appraisals evoke have shift of consistency between the scripts and the referential of the academic appraisals.
Q7. The previous analysis of the ITEF modules is. It in adequacy with the steps of formation from afar.	98%: the respondents estimate that the majority of the educational activities has disciplinary caractère.
Q8. The proposed educational scripts have them of aimed them disciplinary or curriculaires. (Interdisciplinary approach?)	78% the answers confirm the caractère professionnalisant of the contents teaching proposed
Q9. Done the nature of the course and his/her/its séquençement recommend in the multimedia environment she permits the development of the process of professionalization of the academic formation?	82% of the respondents reveal a mobilization uniformisé of resources solicited
Q10. Conceptualization and the modélisation of the contents teaching in the formation do they orient the differential mobilization of expertises from afar?	83%: insufficient contribution limits itself to the level of the modelling of the contents teaching (Absence of the technico-educational engineering differentiated).
Q11. The choices of the numeric resources and the educational activities are sufficient—them to individualize the courses of formation.	95%: the answers show unsuitability between the computer resources and the personalisation of the academic formation courses.

This research appears in the setting of the development research, in engineering of the systems of information. She/it takes like setting of reference several domains, active of the sciences of the education to the engineering of the systems of information. We achieved the conception of the models and tools of educational scenarisation thus [5] F Henri, C Compte, B Charlier—... des technologies en pédagogie ..., 2007 in an interactive conception gait.

Another aspect of this research work concerns the didactic and educational transposition of the methods descended of the domain of the engineering of the needs of the professionals. Our proposition appears therefore in a process of operationalization of the objects of formation survey according to the model profession, whose main objective is to come with the teachers in the conception and the exploitation of educational scripts

(intelligent guidance) and is reuse in other social domain of activity.

We adopted everything along this work a gait project articulating around the phases of analysis of the needs, of conception of the devices of scenarisation, the development of the activities educational support of formation and the structuring of the script adapted to the different profiles of formation.

Of this fact the modelling will be a powerful tool that is going to help us in a doctoral research on the problematic of conceptualization and instrumentalization of the academic formation devices according to the paradigm of the class reversed. The orientations of this research survey are to permit the personalization of the formation courses professionalizing, the follow-up of the adult learner in his/her/its progression of training, in the continuity of his/her/its theoretical formation and

in the autonomy in the construction of the disciplinary and transverse knowledge.

REFERENCES

- [1] Houssaye, M SOËTARD—1991. JSTOR.
- [2] M Macedo-lathe, JM Stairway—Acts of her, 2007—halshs.archives-open.fr.
- [3] V Psyche, OH Mendes—And the Formation. 2003—you-learn.archives-open.fr.
- [4] Philippe jonneart 2002 expertise and socio-constructivism. Paris/Bruxelles: Of Boeck University.
- [5] F Henri, C Compte, B Charlier—of the technologies in pedagogy, 2007—ritpu. “Majoie caBernard, Research and innovation, educational renovation and the technological innovation” (Enkvist, 1994, p. 1).

The instrumentalisation of the MOOCS vector of educational innovation and consecration of the academic training quality

A.T.A. Yalid

Laboratory of Physical Chemistry Analytic, Ben M'Sik Faculty of Sciences, Hassan II University of Casablanca, Casablanca, Morocco

M. Bassiri

Laboratory of Physical Chemistry of Materials, Ben M'Sik Faculty of Sciences, Hassan II University of Casablanca, Casablanca, Morocco

M. Moussted & M. Talbi

Laboratory of Physical Chemistry Analytic, Ben M'Sik Faculty of Sciences, Hassan II University of Casablanca, Casablanca, Morocco

ABSTRACT: The new technologies of information and communication constitute a lever drawing strategic development put forward by the national pact for the industrial emergence. This new orientation aims to sustain the industrial competitiveness of the country. Jointly the system of higher education appears in this perspective, that puts devices of open massive formation in place from a far of which the stake of massification and educational innovation Bernard [Majoie] [1]. Of this fact the quality of the trainings mediatized [Chanier98] [2] depends of the instrumentalisation of this virtual environment.

This numeric revolution constructs the circuit court of basis of a real impact of development of the formation courses professionalizing, that permits the exercise of the profession of the adult learner, the appropriation of expertise technological and the autonomy in the construction of the disciplinary and transverse knowledge [Naymark][3].

In this setting The MOOC is the subject of numerous studies of research that appears mainly in the field of integration of the NTIC again as vector of development of expertise [P mayen.1998a][4]. However, like all human activity these mediatized environments can also be feared under the angle of engineering of the systems of information and formation, especially as he/it modifies a some numbers of the innovating educational and didactic scripts [PAQUETTE, CREVIER & AUBIN 97] [5] in relation to the classic formation from afar (CMOOC and XMOOC).

The object of this survey explorative is to determine a methodological tool of work and organization of the steps from afar educational technical of formation [DIONNE & et al. 99] [6].

This new orientation permits to construct and to conceptualize devices of formation taking in account the requirements of the platforms computerized and the needs of the adult learners [KNOWLES, Malcolm 2005] [7].

He/it appears interesting to underline that the higher education didn't stop renovating and to create his/her/its academic pedagogy in occurrence the emergence of the technical and technological practices of teaching, notably in the domain of the hybrid teaching [8] and of software genius.

In the same way the acceleration of the dynamics of the needs and the development of the mediation tools and of connected generated within the virtual systems of the tools "open source" of which the stake of optimization of computerized training resources.

These activities praxeologique reveal a dynamics of interactive formation (technological mediation pedagogy and intelligent guidance) that contributes to solve the problematic of dwindle of the on line formation and to complete the one of the presently in order to remedy the challenges of the massification and the democratization of access the knowledge, as well as the improvement of the quality of the technological and methodological expertise acquired (to solve the problem of functional illiteracy of the apprenticeship academic).

Keywords: instrumentalisation of the MOOC, educational innovation, Quality of training, engineering andragogique and educational

1 INTRODUCTION

To “learn, it is to become aware and no to follow a way all drawn the best she was”. Jean Piaget

Today, the famous MOOCS (massive open online courses), disembarks to Morocco as allowing the academic students to reach the course in a click, without taking into account the spatio-temporal constraints (synchronous and asynchronous), this new politics educationally will permit to change the paradigm and the methods radically educational support of teaching. This flat shape of educational mediation meet a large success by the techno pedagogue and psych pedagogue like source of incentive and training. The main goal of the on line open a formation is to permit the accessibility and interactivity intelligent collaborative (system of mediation tutorial) [9].

Our objective will have for vocation to put in value of the solutions carried toward a continuous improvement of the design of the environments of virtual training and the modelling of the objects of formation for the processes referentialization of the academic formation courses. Our incentive appears in a dynamics that centers itself more on an anthropological and humanist approach of the learning topics and their previous experience like source of training. These research aim to understand the relations between the features of the learners, the teaching aid and the specificities of the on line training context.

In this setting, the continuous improvement of the technology of the communication and information incited at the universities to move toward a new perspective of instrumentalisation of the systems of information specified to the training that touches the dimension of the autonomy, the responsibility and the independence, however the environment of training plays a fundamental role of which the creation of an educational training space [10] adequate to the young adult learner.

The central question of our research tries to verify the two coins following questions, Q1° quelle correspondance is necessary her to build between the instrumentalisation of the educational design and the quality of the trainings [11] of the learners adult in a MOOC context, Q2° face to the on line formation unhooking, what device of instrumentalisation is necessary it to put in place in order to create the favorable conditions of innovation and educational renovation.

Otherwise the hypothesis of research has the tendency to verify the assumption according to which: the conception and the formalization of the academic formation courses as well as the different fashions of educational engineering of the MOOC devices [12] as vector of innovation. In other words, the recourse to the gait of formation engineering as

tool of piloting efficient of instrumentalisation of the MOOCS appears in the general gait of development of expertise professionalizing.

2 THE REVIEW OF LITERATURE

The setting of reference: who served to the analysis of our survey understands three main measurements:

- Modeling of the systems of information “Modeling profession” [13] in term of structure, obeys some rules as the distribution of the activities, of the powers according to relations and ties that coordinate them.
- the steps of engineering of the formation and educational integrating in the instrumentalisation of an on line formation device.
- the engineering andragogique and educational
- the management by quality total of the formation system.

3 METHODOLOGY PRECONIZED

In relation to the methodology of research we want to adopt a systemic approach [La pointe, 1993] [14] that takes as a basis on the crossing of the contributions descended of the sciences managerial of the quality of the academic trainings and the methods of engineering of the systems of information and formation of the learners on line adult. Of this fact the instrumentalisation of the platform marked and referenced with “educational design”, that permits to have a system of procedures for the development and the implementation of the coherent and reliable (Gustafson and Branch) formation programs (2007) [15] “free translation”.

This methodological gait of research of development and continuous improvement, refer here to the conciliation between the productive activities of the deliverable numeric and the activities of scenarisation encouraging the development of the quality of the formation via MOOC. The retrospective analysis of the activities of instrumentalisation makes itself of way conscientisante and actualizing regrouping two important points. First the analysis of the formation activity that requires the instrumental mediation inciting the inventor to make resorts to a hypothetical interpretation of the objects of study and the activities supports of formation. Then the techno-pedagogue is going to be able to confirm then or to invalidate this interpretation proposed to adapt it according to the context and the population. It is one of the essential roles of conceptualization curricular of the MOOC.

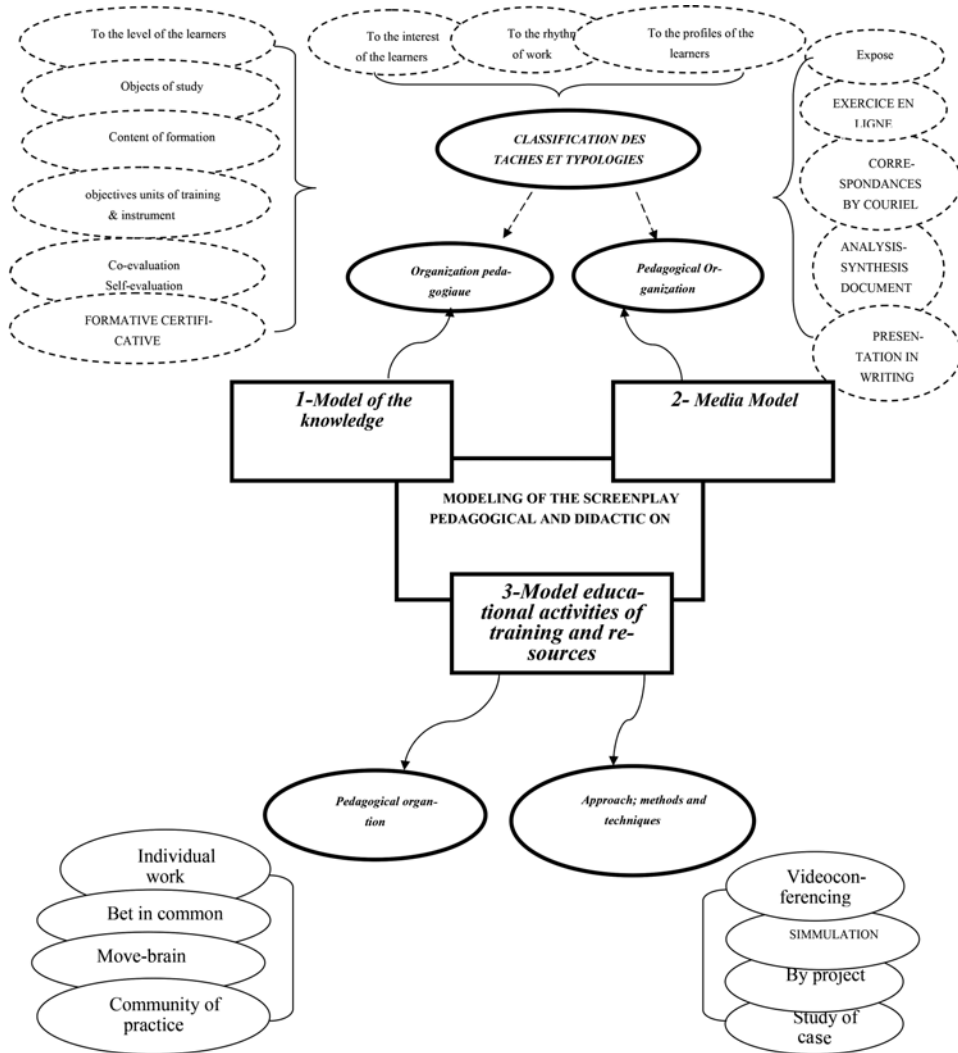


Figure 1. N°1 grid: Instrumentalisation of objects of study scenarios and didactic-pedagogic activity.

4 CONCLUSION

Conceptualization and the instrumentalisation of the MOOC appears in a gait of consecration of the quality of the computerized trainings academic via a platform mediated of type efficient and efficient collaborative. The object of this research survey is to permit the follow-up of the adult learner in his/her/its progression of training, in the continuity of his/her/its theoretical formation and in the enrichment of his's/her's/its's professional expertise's.

The present research is a survey explorative and descriptive of the standards of formalization of the platforms of the open formation from afar

while leaning on a setting of reference (Linard01) [16] that integrates in a coherent global system the actors and their activities, the nature of the contents and the circumstances of the tasks orchestrated, as well as the integration of the technological and educational innovations to create the favorable conditions of investment and engagement in the realization of the different activities of trainings proposed.

He/it appears important to keep to conclude that the instrumentalisation of the MOOC opens an opportunity favorable to the university to create his/her/its pedagogy and to contribute to the quality of the formation courses, to which the educational

and scientific community puts the focal on the NTIC as model of edition of the knowledge.

The MOOC makes an artifact of intelligibility of the offers of devices of formation of the professionals thus in term of communication collaborative, of rich and varied (MondadaPekarek00) interactivity [17], of scénarisation of the contents in real time and in deferred time. He/it centers himself/itself thus on the quality of the trainings professionalizing and on the mission of his/her/its opening on his/her/its socio-economic and cultural environment. It would be about interrogating the nature of the knowledge and the functional expertise's constructed facing a context of demanding more and more competitive work the profitability, the quality and technology.

The main result of this gait of instrumentalisation puts the focal on the hold in account of the modelling and the educational differentiation and of training in the on line formation, of the personalization of the formation courses in the objective to answer the needs of every adult learner and to optimize his/her/its process of training in a perspective social constructivist and interactional learning. The determination of a methodological tool of conception and organization allows us to systematize the interactions and the activities of trainings so that the adult learners.

The instrumentalisation of the computerized environments of training appears as being especially capable to take in account the heterogeneity of the units of formations adapted then to resources of the learners and the diversification of the collective, individual progress personal of training but especially the piloting and the management of this together by the slant of the systems information and process transactional intelligent tutorial for the optimization of the environments virtual of training.

REFERENCES

- [1] Bernard Majoie, «Recherche et innovation, rénovation pédagogique et l'innovation technologique» (Enkvist, 1994, p. 1).
- [2] Chanier, T. (1998). "Hypertexte, hypermédia et apprentissage dans des systèmes d'information et de communication".
- [3] Naymark, J. (2004). "À propos de la publication du FFOD, 'E-learning et knowledge management: quelle convergence' Distances et savoirs," vol. 1, 4, pp. 579–581.
- [4] Mayen P. (1998a) «évolutions dans le parcours professionnel et processus de transformation des compétences». in actes du colloque de la société d'ergonomie de la langue française (SELF). Paris: SELF
- [5] [PAQUETTE, CREVIER & AUBIN 97] Paquette G., Crenier F., Aubin C., *Méthode d'ingénierie d'un système d'apprentissage (MISA)*, Revue Informations In Cognito, numéro 8, 1997.
- [6] [DIONNE & al. 99] Dionne, M., Mercier J., Deschênes A.-J., Bilodeau H., Bourdages L., Gagné P., Lebel C. et Rada-Donath A., *Profil des activités d'encadrement comme soutien à l'apprentissage en formation à distance*, Revue du Conseil Québécois de la Formation à Distance 3, 2, p 69–99, 2003.
- [7] KNOWLES, Malcolm, 2005. *The Adult Learner: The Definitive Classic in Adult Education and Human Resource Development*. Boston; Elsevier.
- [8] Ross, B., Gage, K. Bonk, C., Graham (Éd.) (2006). «Global perspectives on blended learning: Insight from WebCT and our customers in higher education». *The handbook of blended learning: Global perspectives, local designs*, p. 155–168. Pfeiffer, San Francisco, CA.
- [9] Médiation comme «opérateur d'interconnexion entre des séries dont le parcours confère du sens respectivement à ces séries. Elle est clairement, du même coup, une construction de l'analyste au même titre que les séries qu'il identifie» (Simon, 1993, p. 57). Concernant l'objet de notre recherche, la médiation permet de construire un lien entre le champ de l'offre technique et le champ éducatif.
- [10] Giordan A., *Apprendre, comprendre, s'approprier l'environnement*. Cahiers pédagogies, 312, 35–37, 1990.
- [11] Meyer R. (1997). Value-added indicators of school performance: A primer. *Economics of Education Review*, 16 (3).
- [12] Paquette, G. (2002). *L'ingénierie pédagogique: Pour construire l'apprentissage en réseaux*. Sainte-Foy, Canada: Presses de l'Université du Québec.
- [13] Reigeluth, C. M. et Carr-Chellman, A. A. (dir.). (2009). *Instructional-design theories and models: Building a common Knowledge base (vol. III)*. New York, NY: Routledge.
- [14] Lapointe, J. (1993). *L'approche systémique et la technologie de l'éducation*. *Éducatechnologiques*, 1(1).
- [15] Gustafson, K. et Branch, R. (2007). What is instructional design? Dans R. A. Reiser et J. A. Dempsey (eds), *Trends and issues in instructional design and technology (2e éd., p. 11–16)*. Upper Saddle River, NJ: Merrill/Prentice Hall.
- [16] [Linard01] Linard, M. (2001). «Concevoir des environnements pour apprendre: l'activité humaine, cadre organisateur de l'interactivité technique». *Sciences et techniques éducatives (STE)*, vol. 8, pp. 211–238.
- [17] [MondadaPekarek00] Mondada, L. & Pekarek Doehler, S. (2000). «Interaction sociale et cognition située: quels modèles pour la recherche sur l'acquisition des langues?» *Approches interactionnistes de l'acquisition des langues étrangères*, Aile, no 12, pp. 147–174.

Interdisciplinary as a vector of consecration and development metacognitive transversal on line skills

Bassiri Mustapha

*Laboratory of Physical Chemistry of Materials, Ben M'Sik Faculty of Sciences,
Hassan II University of Casablanca, Casablanca, Morocco*

Yalid Amal

*Laboratory of Physical Chemistry Analytic, Ben M'Sik Faculty of Sciences,
Hassan II University of Casablanca, Casablanca, Morocco*

Said Belaaouad & Radid Mohamed

*Laboratory of Physical Chemistry of Materials, Ben M'Sik Faculty of Sciences,
Hassan II University of Casablanca, Casablanca, Morocco*

ABSTRACT: The present tendency of the interdisciplinary paradigm in the societies post—modern, enrolls in a perspective to go beyond virtues recommended a long time by the model of specialization dominant (model of performance). This high level of specialization that was beneficial to the progress of research scientists, but she/it had limitative consequences however to the construction of the borders and partitioning disciplinarians, institutional, and ideological. Delores the reflection bends on the interdisciplinary approach (“meta-disciplinary” or “trans disciplinary” or “multidisciplinary”) that would express this way, extolled by Edgar Morin 1990 [1] according to the paradigm transversally to understand and to surround the complex reality.

We are going to approach the analysis of the choice of our education system Moroccan first while leaning on the sensor interdisciplinary as principle of philosophical conception and organization of the school and academic programs of formation, while trying in a second time to show the difficulties and the limits that hinder his/her/its implementations, his/her/its operationalization and his/her/its assessment in the reality of the educational act.

In the beginning of the years 1990, the system of teaching opted the recourse to the approach transmissive generated by the adoption of the pedagogy by “PPO objective”, answers thus a need to form the specialties and engineers knowing mastered of the well definite stains soliciting the mind pragmatism and analytic that is going to prevail over the fashion of synthetic intuitive, imaginative and global thought.

This report of fact incited the superior commission of the education and the “COSEF formation” in the enhancement of the interdisciplinary importance as powerful vector of development of the disciplinary and transverse expertise, fighting thus against the problematic of the functional illiteracy of the school knowledge (ad hoc Report 1994).

Following this alarming report of the World Bank 1995, the intellectuals and the pedagogues became aware of this art state and looked for tracks of reflection axiological, scientific and praxiological. Philippe Merieu [2] for his/her/its bets in .over. The aspects eminently multi-several references of this trans disciplinary opens proceedings of his/her/its bets in perspective, not only involve the teaching inventors, but also the researchers, the publishers of school and academic works, of the political and institutional leaders who decide the politics educational curriculum.

The present research is a survey of investigating, of retrospective and prospective analysis of the devices and the practices of formation of the professionals of the teachers that can provide to the field of the formation the conceptual bases, susceptible to improve the relevance and the quality of the deliverable.

The object of this survey aims to articulate way very strong two measurements that don't necessarily go together: the theoretical dimension and the operative dimension, that are connected to the gait of the engineering of the devices of formation and Excellency [3] Guy le boterf 2001 and to the contributions of the professional didactics [4] Mayen P. (1998a). (The process of analysis, conceptualization and conduct of the interdisciplinary formation devices). To the bottom, the conditions of the practice of the professional formation let think that he/it is applicable to interest itself/themselves of it to this methodological operative in the goal of location the logics of construction and formalization of the contents teaching organized according to the paradigm of the transversality (capitalization for a real reinvestment in other domains of social formation of activities).

Keywords: transversally; conceptualization; engineering analogical; complexity; professional continuing education; professional didactics and training experiential

1 INTRODUCTION

“I don’t teach anything to my pupils, I tried merely to create the conditions in which they can learn.”
Albert Einstein.

The advent of the new specific educational completions to the secondary education qualifying constitutes without dispute a first in the history of the disciplines of teachings. Indeed it is the first time that they are accompanied with a program describing and organizing the content of teaching in disciplinary and transverse (interdisciplinary) Excellency.

In this perspective a certain number of teacher challenges these new orientations notably in relation to the concept of expertise interdisciplinary that constitutes the corner stone of the formation program. However the introduction of this approach causes numerous questions by the teachers. It is important to underline that this notion of expertise concept is not a new concept, but his/her/its use answers itself more and more in the educational and socio-professional speech. Jointly the use of the principle of transversality can. To be considered like a strategic will (philosophical, epistemological and educational) to solve the problematic of disciplinary carving. He models the conception of a teaching aiming to produce socially controllable Excellency in situations complex, real and unpublished.

In the present context where the matters of teachings try to define their programs of formations and where they have the tendency to pass the simple juxtaposed modules of formation (of the units of trainings). The recourse to a gait of interdisciplinary appears therefore like a powerful strategic foundation [5] Yves based Lenoir 2001 on principles of conception and organization of the contents teaching in order to offer a sort of consistencies internal (school affirmation: school Homomorphism according to the expression of Arnaud stone) and external (social utility of the trainings: school, society).

Otherwise the national charter of the education and the formation since sound before. Subject advances philosophy educational that coins—stretches his/her/its gait of teaching centered on the valorization of the transverse expertise in the formation of the future citizen of tomorrow.

Consequently this new orientation testifies strong and well of a deep vision to an education and formation not only based on the specific disciplinary expertise to very definite stains of training that constitute one end in itself, but

also clearing on transcendent large expertise. She/it answers problematic transversal reinvestable thus beyond the school time justified by the stake social correspondent. This transfer of the apprenticeship school should put in .uvre the means of a training conscientization and actualizing allows the population schooled to manage their socio-professional life in an autonomous and responsible manner. Of this fact a certain numbers of question challenge us: Why and how does interdisciplinary can. It a vector of organization and conception of the contents formation to be? For what stakes? And what are the limits of such approach of conception and organization of the knowledge?

Interdisciplinary should appear in an integrated manner in the didactics of the sciences [6] and would not have to also on no account to overlook the transverse trainings to the detriment of the disciplinary expertise. Delores we are brought to plead for a deep reflection on the four poles that underlie the foundations of interdisciplinary: pole axiological (to enact the finalities the object of transversally: philosophical and epistemological reflection), the scientific pole that founds interdisciplinary, the educational pole (who milked to the pedagogy of devolution and transverse action project would seem more applicable where he/it permits the implication of formed them in conscious and meaning activities to emotional reasoning strong person (the triptych representation, sense, culture) and the pole praxeological of bets in over (implementation).

The fashion of training entry by interdisciplinary can constitute an alternative of solution coveted to limited it of the approach by objective and his/her/its logic of cumulative programming of the units of trainings cut and decontextualizing. Of this fact the didactic treatment would not center itself exclusively on the technical knowledge, decontextualizing and specific to representative situations of definite stains, but oriented also on the nature of the transverse expertise mastered in situations (complex and unforeseeable).

2 THE REVIEW OF LITERATURE

The setting of reference that served to the analysis of our survey consists mainly of the knowledge descended of different fields of development research. We try to observe, by the slant of data provoked and caused by a large bibliographic and experimental survey of the theoretical

conceptualization and the formalization of the practices of the transversality and transfer of training as well as the difficulties that hinder his/her/its operationalization. This research exploratory would develop an understanding and would determine the principles of efficiencies that govern sound the development (inductive gait).

This investigating is oriented by hypotheses of research, that we will propose on the basis of the reports descended of results of empiric studies led par 1° the professional didactics, as well as on a literature carrying on the organizational processes of the engineering of the devices of formation and expertise 2° the didactics of the disciplines. 3° The steps of engineering of formation and expertises. 4° l'analyse strategic and organizational of the education system Moroccan 5° les theoretical perspectives and praxi-ological of the development research. 6° Of the gait of management of the project transversal respecting the plan insurance quality (This procedure has for object to master the gait of realization of the projects of elaborate trans disciplinary collective action to the look of the national strategic orientations of the plan formation—action in all domains of the knowledge).

3 METHODOLOGY PRECONIZED

Between contingency and choice, between description and reconstruction of a methodological setting of conceptualization of a “reflexive return of the scientific topic”, to take the expression of Edgar Morin. This retrospective analysis appears truly in the reflection of Jean Paul Sartre when it says “the man is characterized above all by the overtaking of a situation, by what it succeeds in making what one made it”.

This work of survey and research appears then in a socio-professional orientation: In relation to a positioning exit of my consultant's trajectory—formative in the domain “engineering of formation” and formative teacher to the ENS. These two orientations continue in a third tendency that presents the specificity of my doctoral research works in progress in engineering of the devices of formation and expertise in the course “professionalizing academic students”.

The present survey is a research of investigating, of descriptive and prospective analysis of the national orientations on the concept of interdisciplinary, that can provide to the field of the formation the conceptual bases of construction of the referential of the transverse expertise of formation. She/it offers us an opportunity of modeling of the explanatory theoretical objects of this concept.

The specific question expresses itself of the following manner: facing the difficulties of location

of the logics of construction and formalization of a clear and explicit interdisciplinary approach. What steps of conceptualization of the interdisciplinary objects, of bets in .over, validation and certification? To conceive an engineering of the devices interdisciplinary and to develop expertise transverse metacognitive in a specific context of formation professionalizing, first of all require the shaking of the disciplinary borders.

The object of this research aims to articulate way very strong two measurements that are paradox and antinomy: the theoretical dimension [4] and the operative dimension. The crossing and the complementarity of these two aspects contributes to the setting in relief of a conceptual setting constructed of the key concepts of reference, that encourage in a first time the exchanges between the theoretical foundations. “Epistemological” of the interdisciplinary concept [5] and operative and in a second time that permits to the users to appropriate these instruments and to use them in an adaptive and efficient way.

The objective of this survey exploratory and descriptive was to understand better and to surround the foundations of interdisciplinarity and the processes of his/her/its instrumentalisation in an engineering of formation and expertise's in professionalizing situation. For it, we studied the objects of interdisciplinary studies and the educational and didactic activities of bets in action as well as the modes assessment. For that to make, we tried to plead for the technique of the nominal group as tool of analysis, that allowed us to produce a list important of suggestions: of ideas, of opinions to see solutions to the problematic of interdisciplinarity between the prescribed and the reality what theoretical perspective and methodological gait of organization and of bets in didactic and educational .uvre. To treat this question of departure, we convened 15 professional intervening in the setting of the education and the formation (Three educational supervisors; seven teacher formative to the school normal superior ENS and five teachers of the secondary qualifying representative of the teaching matters: Mathematical. French. EPS. English. Arab), three students of the ENS and two animators (2 doctor ants). Then we examined the suggestions produced to collect the important information, that are classified, listed and priorities according to the order of frequency by individual and collective level-headedness of the statements while choosing those that will become suggestions by the processes of strategic decision.

4 COMMENTARY

The analysis crossed of the existing of the teaching practice and the analysis of the foundations

Table 1. Result of the technical of the nominal group 1. In relation to the institution.

1. Compared to the institution:

To incite a new paradigm of interdisciplinary actions
Powerful lever of struggle against the problematic of the school failure
To remedy the problematic of the amputation of the training on the quantitative and qualitative plan
Encourage the convergence and the harmonization of the interdisciplinary formation programs

2. Philosophical and 2,

Model of inspiration: cognitivist and socio constructivist

3. The object of teaching:

Development of the transverse expertise's

4. Disciplinary reflection:

Internal and external consistency
Vertical and horizontal consistency
Legitimacy of the school knowledge
Construction of an interdisciplinary common identity
Didactization and operationalization of the interdisciplinary transverse expertise
To remedy the carving and disciplinary partitioning and crumbling of the knowledge

5. Objectives of the level taxonomies:

Rest on a structure of the knowledge implying the three domains (cognitive. perceptual—motor and socio affective)
Solicitation of the superior cognitive functions

6. In relation to the conception of the topic:

Powerful vector of acquirement of the objectives of attitudes, method and strategies metacognitive
To palliate the problematic of the parceling of the learning topics
To put the elevated in multidisciplinary contexts
Development at the learner of the cognitive cleverness, the methods and attitudes to solve complex problems
Sources of incentive and sense of training
Vector of incitement at the learner to enroll in the gait of Personal Action Project (PAP)

7. Reflection on the knowledge and the contents teaching:

Contextualization of the knowledge

8. Methods of training and convenient pedagogical:

Reveal a pedagogy centered on the learner
Integrated training developed by activities practice
Based on an active pedagogy valorizing the interests of the learning topics (disciplinary or transverse)
Educational innovation (solicitation of the global intuitive brain of synthesis)
New gait of conception and organization of the teaching content
To give the sense to the trainings school development of the child
Permits the management of the complex and unpublished situations

9. Ergonomics of the stains and situations:

Confronted the learners to real situations of expression and realization

10. Modes of assessments:

Solicitation a high level of procedural resources and metacognitive
The valorization of the training process
Requiring situations of complex and meaning integrations

11. Exigences social and societies:

Utilitarian need of the disciplines
Functionalities of the trainings beyond the school time
Answer to the requirements of the social context (polyvalence and adaptability)

(Continued)

Table 1. (Continued)

12. Limits:

Losses of identity and specificity of some disciplines to the detriment of other
 Utilitarian aim to the service of the requirements of the market globalized
 To incite a new paradigm of interdisciplinary actions
 Powerful lever of struggle against the problematic of the school failure
 To remedy the problematic of the amputation of the training on the quantitative and qualitative plan
 Encourage the convergence and the harmonization of the interdisciplinary formation programs 2. Philosophical and epistemological foundation
 Model of inspiration: cognitiviste and socioconstructiviste 3. the object of teaching
 Development of the transverse expertises 4. Disciplinary reflection
 Internal and external consistency
 Vertical and horizontal consistency
 Legitimacy of the school knowledge
 Construction of an interdisciplinary common identity

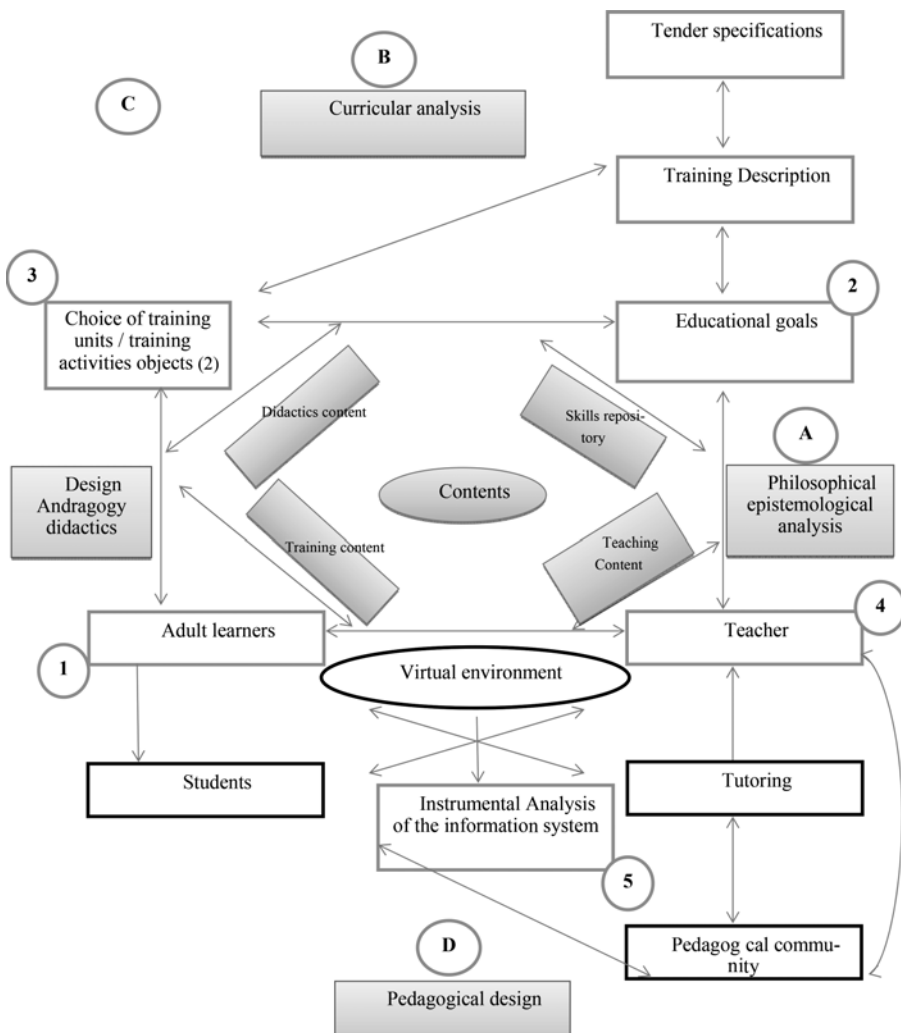


Figure 1.

theoretical and epistemological curriculum shows that the term of the transversality is especially, ambiguous to see utopian between the theoretical speech and the reality of the daily of the class.

The set of the questions examined watch while clear and explicit answer of the concept of transversality and gait of operationalization doesn't exist. In the same way, because of their character polysemous, the proposed representations appear often imperfect and questionable, and few bearers of relevance in tangible didactic solution.

The state of art emerges problems of orders: structural, functional and of fashion faithful thought in favor of the rationalization of the tasks and activities disciplinarian. Our vocation is to build a shape of emblematic teaching organization that puts a new paradigm of interdisciplinary actions in place to remedy the problematic of the amputation of the training on the quantitative and qualitative plan.

The questionnaire specifies that the theme of the transversality affirms itself like vector of scheduling and management of the quality of the trainings and the management of the trans disciplinary expertise. But also she/it suggests steps of disciplinary DE-compartmentalization associated to the notions of process, of interdisciplinary activities, but also of collective projects.

Grid synthesis: Modelling the determinants of the teaching situation interdisciplinary approach.

5 CONCLUSION

To the term of this survey, we can keep to conclude that the emergence and the promotion of the society of the knowledge will make of the interdisciplinary pedagogy a powerful vector of interaction between different disciplines of teachings. He/it distinguishes himself/itself of the approach "Multidisciplinary" that would convene the different disciplines to lead a lot of research, but without real interaction between them while living in partitioned spaces. Interdisciplinary permits to face a specialization and institutional crumbling of the school disciplines of teaching therefore.

This survey considerer interdisciplinary as process of interaction between several disciplines to all levels of construction of the object of training, of the production of the knowledge and expertise reinvestible [7] Julie Thompson Klein 1996. She/it should limit the teaching compartmentalized to the profit of teaching methods that on the one hand encourages work collaborative, and permit to structure the steps metacognitive questioning the behaviorist models of the training again on the other hand: the paradigms of the knowledge

"Rationality, efficiency and efficiency", the reductionist mechanisms of the training and the exclusive knowledge, institutionalized and structured. The debate around interdisciplinary to the academic level [8] N. Rege Colet 2002 puts a term to the partitioning of the disciplines and the reflection on the didactic and educational approaches to clear them. We will generally tempt to define the contours of the concept of "I" interdisciplinary as making resorts the modelling (multidisciplinary, trans disciplinary, multidisciplinary, etc.) that allows us to describe, to surround and to understand his/her/its theoretical foundations, praxiological and educational but the state of art of the formation context to the ENS that explains the failure of numerous interdisciplinary projects made reference adult to the mutual incomprehension of the disciplines (Conceptualization of the terms and concepts of reference outbuildings for mutual comprehension) for example. However the conciliation and the adequacy of specific knowledge and the clean steps amplified the difficulty of determination of the objects of transverse studies.

Facing the lack of transparency, of fear of loss of identity and credibility, incited us to enroll the problematic in an aim of innovation and academic educational renovation while pleading for the couple integration. Diversification far from hegemony "phagocytosis?" Disciplinary.

It is why it is necessary to think upstream on interdisciplinary in term of fashion of thought strategic that a simple relation inter-disciplines. The cumulative aspect would be insufficient to fear this problematic of the functional trainings and constructions of the applicable and original expertise's, transferable in other socio-professional domain of activity. She/it permits to focus the accent on the development of the thought of analysis, reflection and adaptation more deepened of the training based on the relevance of the tools methodological setting up to solve complex and authentic problems.

We recommend like perspective of opening to our research of the approaches of formulation of the principles of transversality according to a horizontal and vertical consistency in exclusive rights of the top or the low of the hierarchy, but rather the fruit of a harmonization between the scheduling strategic curricular and the educational level. The formulation of the objectives of transversality can practices also to the level of the system undertaken, university for the estimable management of the employment and development of the professional expertise's according to an approach strategic mixed curricular of type "«Top-down» et «Botton-up»".

REFERENCES

- [1] Edgar Morin, On interdisciplinarity, Crossroads of the sciences, Acts of the Symposium of the National Committee of Research Scientific Interdisciplinarity, Introduction by François Kourilsky, Editions of the CNRS, Paris, 1990.
- [2] Phellipe Merieu The statute of the pedagogy in the contemporary educational reflection or “To what conditions is the pedagogy able again to hope to change the school?” 1993.
- [3] Guy the boterf to construct expertises individual and collective, 2 edition 2001.
- [4] Mayen P. (1998a). “Evolutions in the professional course and process of transformation of expertises”. In Acts of the symposium of the French-speaking ergonomics Society (SELF). Paris: SELF.
- [5] Yves Lenoir, Bernard Rey, Ivani Fazenda (dir.) The foundations of interdisciplinarity in the formation to the teaching, Sherbrooke, Editions of the CRP 2001.
- [6] J.P. Astofli and M. Develay, The didactics of the sciences. PUF, 1989.
- [7] Julie Thompson Klein. Crossing Boundaries: Knowledge, Disciplinarity, Interdisciplinarity and, Charlottesville VA: University Press of Virginia 1996.
- [8] N. Rege Colet academic Teaching and interdisciplinarity, Of Boeck-Wesmael, 2002.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Interactive formation from afar and cognitive intelligence increased at the adult learners: Case of the master technological engineering for the education and the formation

Bassiri Mustapha

*Laboratory of Physical Chemistry of Materials, Ben M'Sik Faculty of Sciences,
Hassan II University of Casablanca, Casablanca, Morocco*

Yalid Amal

*Laboratory of Physical Chemistry Analytic, Ben M'Sik Faculty of Sciences,
Hassan II University of Casablanca, Casablanca, Morocco*

Said Belaouad & Radid Mohamed

*Laboratory of Physical Chemistry of Materials, Ben M'Sik Faculty of Sciences,
Hassan II University of Casablanca, Casablanca, Morocco*

ABSTRACT: This survey has been achieved during a research of action led by the academic students of the faculty of the sciences BEN M'SIK of Casablanca within the physical chemistry laboratory of the Materials (LCPM). We chose the studying Master ITEF "Engineering and Technologies for the education and the Formation" as corpus of empiric sample of our research to discover, to construct our object of study and to enhance the empiric illustrations to present in the setting of the problematic of research ("survey of case" that we granted to the survey of the cognitive functions of the thought of analysis meta-cognitive and creative reasoning of the exponential intelligence.

Of point of epistemological view: This research is located in the field of development research. His/her/its goal: is to explain the tie of reason to effect between the profiles of the learners and the conditions of teachings presiding to the process of the cognitive expertise's high-level according to the taxonomy of Bloom revisited by Anderson and Krathwohl (2001).

The present research is also said descriptive insofar as it permits: 1) the location logical of the objects of study to finalities metacognitive 2) the Generality and transfer of the development metacognitive principles in other socio-professional domain 3) the reference to a systematic device of description of the process solicited OF Ketele, J.-M., Roegiers, X. (1996). The survey tries to put in inscription the conditions of validity of the knowledge metacognitive and construction of expertise's conscientization Piaget, J. (1974) and actualizing OF IT. Bertrand (1998). It is necessary to recognize that the mobilization of the cognitive resources high-level, require on behalf of the inventor of the programs to take the conscientization of the solicitation activities not only, but also their mastery in an efficient manner in real, complex and varied situations of training.

The transformation of the generation Z of the young academic adult learners that needs to mark a rupture with the shapes of mechanization of the thought acquired during the cycle of license formation toward models of training of the creativeness, the autonomous and the liberty in the decision making. It is in this dynamics of understanding of the steps metacognitive that our survey of research appears, while leaning on the reports of the bibliographic studies that showed the main problem of the development of the superior cognitive functions in an environment of interactive virtual mediation. The activation of this process is not the lack of the knowledge acquired among the students, but to the inability of the choices of the technical-educational devices adapted to the different styles of trainings of the adult learners Kolb, d.a. 1985 and their interactions to the processes of mobilization of the steps metacognitive: structuring of the cognitive, anticipatory networks, of scheduling and control "feedback" of the programs of proceduralization of the knowledge. The problem becomes attached to the register of the steps of conception of the varied devices of formation and varied granting a pre-eminence to the intellectualization of the on line formation. The present research of development titled "Formation from afar interactive and artificial intelligence increased at the adult learners" is a descriptive research of mixed type, conjugating qualitative and quantitative data at a time. The specific question of research expresses itself of the following

manner: What correspondences to establish between the analysis of the variables omen “style of training” and program “contained of formation” and the development of the strategies metacognitive, at adult learners, in situation of training in a specific context of on line formation? The questions and the hypotheses of research that have been recommended are composed of two aspect complementary. The first expresses the quantitative data collected by the studying Master ITEFS. The second aspect regroups questionnaires of analyses and identification of the training styles: the Learning Style Questionnaire, French version abridged (LSQ-F) of Small fort, Goatherd, The Berge, Leblanc and Amyot (2000) and of the measurements of the processes and knowledge put in play at the time of the process teaching training Bloom, b. 1969 revisited Anderson and krathwohd. So much what concerns the qualitative data have been collected by a sample of 25 respondents, following an interview semi controlled putting the accent on the one hand on the fashions educational preferential (concrete experience or reflexive experience or abstract conceptualization or active experimentation) respondents to their needs and on the other hand to the solicitation of a high level of mobilization of the cognitive resources.

Keywords: e-learning; styles of trainings; metacognition; pedagogy; andragogy; training experiential of Kolb; course of training differentiated

1 INTRODUCTION

The prestigious development of the society of the knowledge, information and the communication makes of the formation a preoccupation adult of the daily of numerous formations of adults. We progress in a society where the training appears “all along life” lifelong “Learning” rather than in preparation of the existence [1]. (Merchant, 1997; Bernatchez 2000).

Morocco appears in this quest of the quality of the formation, while considering the education and the formation as national priority after the territorial integrity (national charter of the education and the formation 1999). He/it is to note that the ultimate objective of this strategy educational is “1” optimization of the employment of the educational numeric resources in order to make the best use of the new technologies of information and communication. To this consideration it is important to recall and to underline that the rigorous and efficient integration of these new technologies in the process of initial formation must overlook on no account the measurements did the reflexive training recommended by the adoption of the approach by “APC expertise”. This new orientation aspires to make to advance and to improve the quality of the teaching, on the plans of the programs of formations and methods technical—educational adapted to the environment of mediations computerized.

The gait of research is centered on two problematic: The first makes references adult to the assessment of the cognitive cleverness and their report to the referential of expertise’s aimed in the descriptive of formation modular “engineering of formation and expertise’s”. The problematic second considers spreading research subsequently to the problems of the formation self-culture based on the paradigm of creativeness like a tool of piloting efficient of innovation and renovation and that

provides to the fields of the academic teaching the reflexive and creative parameters of conceptual and methodological modelling of the teaching.

We try to observe, by the slant of data provoked and caused by a large survey and research bibliographic exploratory to construct the setting of reference for the analysis of the data. We go all along our survey adopted the process of Anderson and Krathwod and Kolb that go permits us to test and to remedy the hypotheses to explain the process of working metacognitive in an explicit manner. The methodological process is founded on the direct and indirect (inferred from the sessions of justification and debriefing lived during the degree course of formation according to the approach of intellectualization of the discipline) observation.

The proceduralization of the research protocol organizes itself in two times. First of all, we analyzed the cognitive cleverness mobilized during different sessions of animations to the module of engineering of formation of expertise’s and to compare these cleverness acquired with those to solicit in other units of teaching. The setting of reference takes pushes the grid of analysis of “Anderson and krathwohd” taxonomy of “Bloom revisited” concerning the training of the cognitive cleverness high-level and that carries on two objects of study: 1°) to identify the features of the measurements of the “cognitive process” the modeling of the “cognitive process dimension” (Remember. Understand. Apply. Analysis. Evaluate—Create.) and 2°) to Determine the measurements of knowledge solicited “the knowledge dimension” and (Factual knowledge—Conceptual knowledge—Procedural knowledge—Meta-cognitive knowledge) are not always cognitive cleverness (to listen to, to say, etc.) in the sense of Anderson and Krathwohl (2001). A posteriori we are going to plead to the analysis of the activities put in .uvre by the studying Master “look critiques on the different obligatory readings

transmitted on line” and the situation of assessment and final integration of the trainings “Assessment summative”. It allowed us to identify the evolution of different measurements bound to the reflexive analysis of the situations of animation, to the auto evaluation of his/her/its own intervention “self-knowledge” and to the assessment of the variables: centration—attention. And decision making in relation to problematic landed. Then, we analyzed the set of the different meadow production—projects presented in the assessment formative of mid-course, in the goal to determine the factors that condition the formalization of the practices metacognitive to answer the requirements of the referential of expertise’s eligible metacognitive.

The specific question of research expresses itself of the following manner: What is the nature of the reports from afar to establish between interactive educational practice and cognitive intelligence increased, at adult learners, in situation of training in a specific context of formation?

The relative under-questions’ to research:

- Q1: Does a dominant training style, at adult learners, in situation of training, in a specific context of on line formation permitting the development preferential metacognitive, exist?
- Q2: What correspondences to observe between educational practice and the development metacognitive of adult learners, in a specific context of formation?
- Q3: How to characterize the approaches and the steps of formation to enroll in the paradigm metacognitive, and this in a specific context of on line formation?

So if one subscribes to the hypothesis that the practices educational styles of training of the adult learners, express the principles and the fundamental mechanisms that can fill the mission with development of expertise’s cognitive strategies of superior level in a computerized training environment.

2 THE REVIEW OF LITERATURE

The theoretical foundations of the concept of metacognition indicate that the determinants of the functions metacognitive play a crucial role in the dynamics of constructions of the cognitive expertise’s according to the model of the autonomy and the reflexivity of the action and where the adult learner splicing that topic proactive learns gradual manner of the knowledge acquired and knowledge of his/her/its own cognitive processes [2] (Flavell 1979). In the case of the analysis of the situations and activities of formation we could show that the variable measurements of the knowledge and

measurements of the cognitive processes enter in account in the assessment of the levels the engagement cognitive of the students, their feeling of efficiency [3] Bandura 2003 and their performance cognitive would permit to open other ways of intervention that the deletion of the exercise.

In this research exploratory, we led a survey of analysis of the efficient activities of formation while trying to conceive and to drive devices of formation permitting the development strategies metacognitive “to develop an increased intelligence”. The analysis of the questionings semi-directive permitted us to collect the encoded data as well as the representations of the learners adult “structure of welcome” with regard to the approaches of trainings and expertise’s waited while trying to value their degree of interrelationship with the object of development of the knowledge and expertise’s metacognitive.

In the same way the analysis of the conception of the projects of mid-course also reveals us the measure the doorstep of the cognitive working via the analysis of the plans and the methodological steps of conceptualizations of the meadow-projects of formation and the organization of the seminary “Assessment of the academic formation system: Reality and perspective”.

A first level of reflection is provided by the researchers who position and infers the analysis of the reasoning of the process cognitive of the thought from the survey of the educational activities, of the steps of teachings, of the impact of the intelligent guidance, of the identification of the training styles, of the modes of interaction and communication and the instrumental theories, of industrialization, connection and independence and autonomy of formation in e-learning.

The setting of reference that served under-abstractly to the survey of the theoretical of our research president to the definition of the main hypotheses of work is located in the crossing of the domains of the differential psychology, of the cognitive psychology. And socio constructivist [4] Vygotsky, L.S. (1984). But one finds the traces of the personality’s former psychology (cognitive style) there also.

The performance of the on line trainings is inferred from the quality of solicitation of the process metacognitive and knowledge constructed underlying. The instrumental constraints of this environment are bound on the one hand to the virtual formation device, to the nature of the situations of trainings, to the specificity of the educational activities proposed and to the present resources of the adult learners (their ways of which they use the knowledge and of the way of which they discern their faculties). The cognitive fashions of working of the adult learners are a matter

therefore notably for the exam of the steps metacognitive, to the measurements knowledge that the topic to his/her/its own cognitive processes [5] Flavell, J.H. (1977) and to the processes of self-regulation control his/her/its activity [6] Brown, HAS.L 1987 by the setting in game of the mechanism of awareness, proceduralization and construction of sense to encourage the conditions of a real training of the Metacognitive strategies (the active and self-regulating dimension of the training).

In the survey of our case the process metacognitive will be to analyze in relation to the systems of knowledge méta-knowledge concern the requirements of the objects of study and the educational activities of the instrumental tasks, the strategies setting up by the topic learning, of his/her/its way to proceed and to his/her/its styles of trainings. The object of research also concerns the self-regulation of the activity before, during and after an intervention of course.

3 METHODOLOGY PRECONIZED

Our gait of work is inspired of the method of research development of the tools of conceptualization of the reflexive formation devices. She/it tries to assure and to rationalize the construction of expertise's vector metacognitive of development of the superior cognitive functions in a context of formation hybridizes. The recommended research-action implies the epistemological responsibility splicing teacher researcher and the respect of the deontology of the educational practices that aim ethics and the validity at a time [7] Narcy-Combes, J-P. 2005. Offset 2008 of the

experimentation. This survey has for object the conceptualization of a design Anagogical and didactical-educational of formation of the adults in context of training from afar, while trying to enroll in the new epistemological paradigm of the knowledge metacognitive.

Consequently the problem of understanding of the working of the thought mechanisms: To question this capacity of working of the cognitive process (or mental acts) in relation to the measurements of knowledge (the knowledge dimension) and measurements of the cognitive process (the cognitive process dimension) bets in relief in the process of acquirement of the knowledge at the adult learners. This problematic will be treated in two times: We go shown first how operationalize and to symbolize the concept of knowledge metacognitive by the slant of the organization of the objects of studies and the educational activities and to put it from afar in relation with the instrumental theories of the formation. In second time we will evoke the gait of development of a conceptual model of the content—program and of the procedures of formation and assessment answering the foundation of the paradigm metacognitive.

Context of analysis: facing the problematic of the on line formation unhooking in report to the styles of training of the adults and the results of microscopic analysis of different didactic educational activities that showed some limits in this device of formation (solicitation of a lower level in the taxonomy of bloom: memorization—understanding and application).

The definition of the metacognition is complexities by his/her/its character polysemic of

Table 1. The grid of analysis recommended: Anderson and krathwohd “raining high-level cognitive”.

The cognitive process dimension						
The knowledge dimension	Remember	Understand	Apply	Analyze	Evaluate	Create
Factual knowledge	25%–22%	23%	28%	16%	10%	
Conceptual knowledge	38%					
Procedural knowledge	22%					
Meta-cognitive knowledge	15%					

Taxonomy revisited of bloom.

Table 2. Grid of analysis and identification of the styles of training of the studying Master ITEF.

Preferential fashion of on line formation	Answer of the studying NR 26
Experience concrete “Accommodator”	16
Divergent “reflexive experience”	05
Assimilative “abstract conceptualization”	02
Convergent “active experimentation”	03

ITEF: Technologies for the education and the formation.
NR: Number.

Table 3. Grid of analysis of the approaches—Steps & educational methods in relation with the cognitive functions.

Approaches—Steps & educational methods	Percentages
Approaches 100%	Trans missive 45%
	Behaviorist 26%
	Constructivist 15%
	Socio constructivist 14%
Demarches & strategies 100%	Inductive 22%
	Deductive 68%
	Dialectic 10%
Methods pedagogies	Demonstrative 23%
	Analogue 05%
	Magistrat 42%
	Interrogative 14%
	Discovered 16%

Table 4. Controlled questionnaire 3 semi: identification of the training styles semi controlled: identification of the training styles.

Questionnement in the studying master ITEFS	Answers
Q1. Do the courses of Master formation in the conception and the determination of the formation objectifs take—them in account your needs, your waiting and your preoccupations of formation professionalizing?	90%: some respondents estimate that the objectifs is exclusively prescrit in accordance with those to recommend in the special prescription notebook
Q2. The language of teaching of the different modules is—it in consistency with the linguistic contextualisation the tasks of trainings recommended	45% of the answers mentionées generally a shift of specific semantic order
Q3. Does the educational mediation encourage t— it the solicitation of the superior cognitive functions?	76% of the students reveal the rigorous difficulties of accompaniment soliciting the metaconnaissances
Q4. Do news and the orders of realizations of the tasks of trainings prescrites permit. them the mobilizations of the process metacognitif?	78% of the learners think that the majority of the stains propsees is defined mobilizing processes of understanding and application/and of the factual and conceptual knowledge
Q5. Do the definition and the discussions of the formation objectives through the sessions of justifications and debriefing encourage—them of the returns reflexive and of the formation proactive strategies?	89% of the respondents explain that the two steps of formation of the adults are to recommend in two modules of formation on eight
Q6. The report to the knowledge detrmine t .il an environment of dialogue, of proceedings of idea and conflict socicognitif of explicit training?	65% of the answers estimate that the report to the determined knowledge again the report to the power “relation educational hiarchique!”
Q7. Does the previous analysis of the ITEF modules recommend it of the steps of teaching of pocéduralisation and reflecting abstraction?	78% evoke a shift between the steps prescrites in the notebook of the charges and the educational practice
Q8. The ethical setting and the contract didactics setting up permits them to tie intercourse of collaboration interactive “guidance intelligent” pledge dedeveloppement of the capability of auto and heterorégulation conscientisantes?	82% confirm the lack of a clear and explicit didactic contract where him ya engagement and mutual respect of the different taking parts
Q9. Done the nature of the course and his/her/its séquencement recommend in the multimedia environment she to write you down in the paradigm métacognitif permits?	83%: insufficient contribution limits itself to the level of the discussions in round table in présentiel (Absence of the guidance educational community—formed)
Q10. The conception and the structure of the contents teaching in the formation does the orient the solicitation of the cognitive ressources in an optimal manner from afar?	95%: some answers are negative (model informationel without report with an engineering of the needs)
Q11. The choices of resources and the educational activities and training are sufficient—them to create a real reflexive training in this computerized training environment.	98%: lack of the activities dynamisantes of the trainings, to know the development of the métacognition and the teaching of strategies reflexive

this concept. He/it dealt with us discriminating to make resort to the intersection of three models: 1° Model of Anderson and Krathwohd 2°). Model of Flavell” “the knowledge that one has his/her/its own cognitive processes, their products and everything that there touches” and 3°) the dynamic model of Nelson and Narrens 1990 [8], that has for advantage to center itself/themselves on the processes put in play (process of assessment and control of the cognitive processes) and that represent the set of the mental operations on the mental operations and no on the thought contents. For example, at the time of the interviews and questioning semi-directif of the adult learners about the program of formation Master ITEF, the main objective doesn't limit itself to identify their structural representations: together of the information judged useful to land and to treat a discipline of teaching. We describe the behavior observed of the learners according to the process of the mental operations achieved on the object to know “program curricular then”. On the other hand, if we wish to enroll in this paradigm metacognitive of Nelson, the learners adults “object of our research” should be challenged to analyze their own benefit and individual and collective realization. It is for it that we oriented our assumption of research on the analysis of the strategies adopted to value and proceduralize their meadow project of formation mid-course during the sessions of justifications and final debriefing soliciting a reflexive return on investment on experience while focusing on the future actions of remediation, management and piloting of their projects of research. It will be then about processes of real development of the knowledge metacognitive because the mental operation is exercised on their own mental operations.

The finality of intervention aims to put in inscription at least on two axes of reflection. On the one hand, she/it addresses the teachers, inventors, guardians and professors who try to underline the necessity to elaborate directly and to offer some varied courses of formation. But she/it also addresses the adult learners as factor of incentive through a better understanding of their own mechanisms of training, and as factor of optimization of the training steps metacognitive “methods of <procedural zing> of the constructed expertise's”.

4 CONCLUSION

The development metacognitive is the main motor of the construction of knowledge and expertise's metacognitive. This process permits the improvement of the reflexive observation and the conceptual and methodological modelling of the reasoning and hold of decisions to solve complex and unpub-

lished problems. The object of the reflection explicit metacognitive causes the awareness of the permanently the adult learner splicing that “capable topic” before being formed object. Orientation toward formation of a capable topic is going to orient us toward the change of the epistemological paradigm, reversing the report to the knowledge and to the activities of production of these knowledge. This activity to incite the learner topic to be social actor responsible for his/her/its development, to be capable to describe his/her/its cognitive strategies put in relief to solve the problems and to have the faculty of reconfigure his/her/its representations and his/her/its cognitive resources to formalize projects of actions actualized efficient and applicable (analysis, identification and to anticipate the indicators and of the conditions of realization of the stains asked and to discover their implications of it in other social domain of activities).

The domain of the higher education, the understanding of the working of the knowledge structural metacognitive on the processes of thoughts self-reflexive (to the sense neuropsychological) and on the contents thought hetero evaluation (to the sense cognitive—behaviorist), exploring the capacities of the adult learners to understand their own mental phenomena and those of the other students in a perspective of transfer and reuse in other social domain of activities. The development of the metacognition according to the perspective socio-constructivist and interactive remains tributary of the conception of the objects of studies and activities educational support of teaching.

In this work we restore the different measurements of the knowledge and cognitive process solicited during the sessions of Master formations while taking the Anderson model and krathwohd as setting of priority reference. This model of analysis puts in relief the complementary crossing enters two aspects of the working of the thought. He/it appears important to keep, that it is therefore discriminating to focus the reflection on the educational practices putting splicing that vector of consecration and development of a state of cognitive balance of consciousness and systematic and adaptive self-regulation. This survey of research explored by our academic students allows us to deduct that it is also possible to instrumentalist and to formalize objects of training e-learning, developing these cognitive measurements high-level in a very meaningful manner. The gotten results open a reflection deepened on the engineering operational Anagogical of the metacognitive consciousness in a computerized environment (conceptualization and the formalization of the didactic and educational objects) therefore. To conclude the modes of observation diagnosis overtly achieved by the students who

are been called to judge and to describe by a free and faithful verbalization the preferential situations guaranteeing the quality of the formation metacognitive. In the beginning of research the students show behaviors of fear and anxiety facing this situation of new formation, that doesn't generally answer their needs and do valorize neither their experiences nor their personal action project [9] Meirieu, P. (1993). The analysis of the educational objects of formation omits the dimension metacognitive (who represents the way whose adult learners represent, conceptualize and procedurally their trainings) from afar while putting in inscription the valorization of the conceptual and procedural knowledge and the dimension of the creative cognitive process that will mark a rupture paradigmatic of the conception of the specific traditional activities of memorization, understanding and application to stains defined "The system: Goal" (Operation to put in .over and criteria's of successes is defined well to the previous). The awareness in this virtual environment becomes a powerful lever of autonomy, independence and self-knowledge permitting to know itself/themselves better and better to interact within an educational community. The phase of conception of the didactic and educational engineering gait in formation lines off should put the accent on the setting up of an educational instrument that aims the training in computerized environment capable to evolve to the different styles of trainings to answer the problematic of the development of the courses professionalizing individualized and the adaptation of the training in e-learning (of which the

content of teaching, the exercises, the orders of work, modes of assessments etc.).

REFERENCES

- [1] Merchant, L. (2000) characteristic and problematic specific to the academic formation by videoconference and telematics In *Cyberspace and Open Formations. Toward a mutation of the formation practices? Under the directions of Alava, S* (2000).
- [2] Flavell, J.H. (1979). Métacognition and cognitive monitoring. To new area of cognitive developmental inquiry. *American Psychologist*, 3, 906-911.
- [3] Bandura, TO. (2003). Auto-efficiency. The feeling of personal efficiency (Trad. J. Lecomte). Brussels: Of Boeck. (Original published in 1997).
- [4] Vygotsky, L.S. (1978) *Mind Society* in. Cambridge: Harvard University Press.
- [5] Flavell J.H. W and H.M., *Metamemory, Perspective in one the Development Memory of and Cognition*, R.V. Kail W and Hagen, Editors. 1977, Erlbaum: Hillsdale, N.J. P. 3.33.
- [6] Brown, HAS.L., *Metacognition, incentive, and understanding, Metacognition in, ministerial control, self-regulation and other more mysterious mechanisms*, F.E. Weinert R and Kluwe, Editors. 1987, Lawrence Erlbaum Associâtes: Hillsdale. p. 65.116.
- [7] Narcy-Combes, J-P. 2005. Ofsted 2008. The changing landscape of languages." HMI070053. www.ofsted.gov.uk/publications/07005318.
- [8] Nelson, T.Oh. L and Narens, the psychology of learning and incentive, *Metamemory in: To theoretical framework and new findings*. 1990, Academic Press: New York 19.
- [9] Meirieu, P. (1993). *To learn... yes, but how?* Paris: ESF, Coll. Pedagogies.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

The engineering andragogical in the device “blended learning”: Case of the academic formation courses professionalizing “development of a conceptual and methodological” setting

Bassiri Mustapha

*Laboratory of Physical Chemistry of Materials, Ben M'Sik Faculty of Sciences,
Hassan II University of Casablanca, Casablanca, Morocco*

Yalid Amal

*Laboratory of Physical Chemistry Analytic, Ben M'Sik Faculty of Sciences,
Hassan II University of Casablanca, Casablanca, Morocco*

Said Belaouad & Radid Mohamed

*Laboratory of Physical Chemistry of Materials, Ben M'Sik Faculty of Sciences,
Hassan II University of Casablanca, Casablanca, Morocco*

ABSTRACT: The system of higher education didn't stop innovating and to renovate his/her/its academic pedagogy in the stake fundamental of development of the professional expertise's aiming to answer the needs of a society in constant mutation, to the difficulties of the job market, to the instability of the professional careers and to the increasing complexity of the professions that requires a high level of technification and automation. The involvement of the new technologies of information and NTIC communication in this national strategy arena to for objectives to fight against unemployment and to improve the competitiveness of the technological market globalized. This new orientation, bring us to think on the restructuring and the requalification of the professionalizing formation splicing that vector of development socio-economic Wittorski R., Ardouin T. (2012).

In the daily of the class, the professional academic formation of the young adults leash to think that it is applicable to interest itself/themselves from afar of it to the methodological operative Andragogical via the hybrid teaching in the goal of location of the logics of construction of the objects of teachings and formalization of the educational scripts corresponds. The context of research that is ours, to know the training in blended learning, allows us as a matter of course to interest us to the student as young learner adult “concept of learning”, because this last constitutes the majority of the public schooled registered in the degree course of full-time academic formation in order to get a complement of initial formation. Within sight of what preceded the orientation of the engineering anagogical in device “Learning blended has for objective to learn to learn” has to has according to the paradigm metacognitive Anderson, L. W. Krathwohl and, D. R., and al (Ed.) (2001), that is to acquire the technological expertise's that constitute a powerful springboard for the development of the methodological, strategic expertise's and communicational wagers a teaching professionalizing, rather than to learn conceptual knowledge (exclusive sollicitation of the mechanisms of memorization and understanding according to the taxonomy of bloom).

The central question of our research expresses itself of the following manner: facing the difficulties of location of the standards of conceptualization and formalization of an engineering Andragogical in the hybrid teaching. What methodological gait is necessary—it to put in over to promote the courses academic professionalizing? and rationalization of the professionalizing academic formation, encouraging the improvement thus their capacities to face and to solve complex and unpublished problems.

Two questionnaires have been submitted to the respondents: The first questionnaire intended to the students: a specific sample of studying them of the faculty of the Ben sciences me Casablanca sik (variable omen), organizing itself/themselves author of the following data: 1°) Their features (domain of survey; demographic features; social and professional and the technical cleverness); 2°) Their needs and their waiting in the comparing to the objectives of formations recommended in the notebook of prescription special CPS; 3°) Identification of their training “styles the Learning Style Questionnaire, French version abridged (LSQ-F) of Small fort, Goatherd, Théberge, Leblanc and Amyot (2000)”.

The second questionnaire intended to the teachers (Variable program and product) to analyze the perceptions of the respondents as for their own conception of the virtual formation” devices the objectives of formations—the content, the didactical-educational means, modes of assessment recommended” to put in inscription the models and the standards of formation adapted Me to the context of the Ben faculty sik of Casablanca (survey of feasibility). Within sight of what precedes us proceeded to a gait of analysis of the latent program while only keeping the units of training bearer of senses and meaningful information in relation to the central question of research for the determination of the criteria’s: 1) the perception of the respondents as for their own style of training and 2) the sense given by the respondents to their own situation of on line training. The analysis of content has been kept like method of work for the treatment of the data qualitative.

Keywords: Engineering anagogical; Blended Learning; education of the adults; educational engineering; professional continuing education; professional didactics and training experiential

1 INTRODUCTION

“The on line training constitutes a catalyst that redoes the picture of the education, join in network the academic establishments, answers the governmental orientations and oblige the professor to think about his/her/its new role” L’autrice

The universities Moroccans appear in this orientation while becoming a space privileged of education to the values of citizenship and civic behavior, while granting a fundamental role to the construction of expertise’s technological, cultural and methodological pledge of professionalization of the student’s profession “national charter of education and formation 1999”. Even though the teaching in presental constitutes the fashion the more used in the universities, the on line training brings us to establish a new change of paradigm of the report to the knowledge “virtualization of the knowledge”. The faculty of the sciences well Me sik of Casablanca bestowed a project of integration and promotion of the NTIC support of teaching and scientific research. Jointly the assessment of the descriptive of formation and the different projects of research shows strong and well the importance granted to the engineering of the technological systems for the education and the formation that gave birth to the emergence of the “society of information” as powerful determinant of the economic and scientific progress (Glikman 2002).

Leave elsewhere the growth important of the strengths of the academic students, the demand of the job market in technological expertise’s, the preparation of the students to a learning society and the requirements of the new approaches curricular that grant a fundamental role to the faculties of analyses, of reflection and to can of adapted to the technical and technological context “the diversity of the virtual environments”, make that the traditional fashions of teaching in presently, appear henceforth insufficient for the appropriation and the construction of the disciplinary and

transverse expertise. In this setting the arrival of the technologies of information and the communication made appear the place and the relevance of the new on line formation devices like one fashion of teaching to the first plan in the domain of the education and the training (Merriam and Caffarella 1991).

The question of the rationalization and the optimization of the teaching blended learning is to put directly in relation with the problematic of the teaching of the adults in academic formation course professionalizing, to this type of formation to know the problem of massification and abundance of formation in presently.

Us survey of investigating appears in the process of research development of the quality of the programs of training of the adult learners in u hybrid teaching. The project of integration of the educational practices of alternation “presently formation and from afar” in the degree course of initial and continuous formation enrolls in the paradigm of the innovation and creativeness. Of this fact the analysis of the factors that influences the conception of the objects, the educational activities and the modes of assessment of the devices remained tributary of an engineering simultaneous andragogical mobilizing manner harmonious and integrated the different determinants of the situation of hybrid teaching (Students. Teacher in presently. Virtual educational Community. on line intelligent Guardian—Referential of expertise’s. Nature of the Knowledge. Specificity of the formation context).

This approach of the research-action will first make the accent on the analysis of the state of art of the strong points, and of the opportunities offered by the hybrid teaching, while insisting on the measure of impact and the effects of these new technologies, on the assessment of the feasibility of their integration in the academic educational scripts setting in .overs. Jointly this research will permit the development of the devices of engineering andragogical and expertise’s in this

environment Blended Learning, articulating the contents formation and the needs of the learner's adult goshawks of the strategies of collaborations intelligent tutorials. This vision of assessment of the norms and standards that presiding to the establishment of the methodology of instruments and mixed most tangible formalization of the devices answers a problematic of innovation and quality of the academic trainings. We will be brought therefore to conceptualization and to the setting up of a systemic modelling of the objects of teaching and activities educational support of training, articulating the technological aspects, the aspects psycho educational, the methodological aspects and organizational of the hybrid context, without omitting the aspects of the setting of referential reference" of the eligible national expertise's" and the features of the learners adult actors of the auto-formation process.

The recourse to a magazine of the scientific literature of the mixed teaching, allowed us to illuminate the standards of development and the steps opened out in the process of formalization of these devices. We kept therefore for our survey of research the focusing on the teaching Blended learning splicing that technical device, model of design instrumentalisation educational curricular and processes of organization of the didactic scripts and activities of teaching—training that has for crucial goal to provide to the fields of the education and the formation of the adults the favorable conditions of the management of the courses of auto formation and construction of the professional expertise's. This fashion of hybrid formation requires the conciliation of the technical requirements of the virtual environment "leaning on an electronic technology" and the educational requirements of conception of the computerized contents (module-based training) and pedagogical scenario of the educational activities.

The approach of the mixed formation founds on the conception of the environments of training susceptible to answer the principles of continuity, extension and complementarity of the teaching presently first of all and in a second place makes call in a cascade of stages of analyses systemic of the components: Technological 1°—components of the computer resources 2°—profiles and needs of the adult learner placed in the center of the on line training system and that makes of him a dynamic and conscious actor of his/her/its training process, 3°—The approach socio constructivist and cognitivist of the training of the institutional vision 4°—The theories of the systems of communication and interaction sheltered by the virtual environment [1] (Golden 1998). If we leave from the assumption that the formation Learning blended belongs to the category of the teaching,

that organizes the relation presumed from afar between the teaching presently and teaching to end to propose courses of training professionalizing varied and adapted in the goal of optimization of the process of training of the academic courses.

2 THE REVIEW OF LITERATURE

The setting that will serve to analyze the relation presumed between the engineering of formation andragogy and the professionalization of the courses of training of the students' academic in teach it Blended Learning corresponds to the definition of the setting of reference that we elaborated from the key concepts descended of different fields of research. These concepts that also served to the recession of the writings and to the definition of the conceptual and methodological setting of this survey exploratory. We identified the following domains thus: 1) the training in the formations hybrid and 2) the engineering of the formation of the adults 3) the contributions of the professional didactics (professionalization of the formation courses) 4) the theories of the sciences of the education.

The analysis of the literature shows besides that one can distinguish two approaches of research mainly: The first approach searches for brings itself to the descriptive analysis that consists in analyzing the main theories of training of the adults and to determine the essence of the foundations theories orchestrate transferable in the action and the act educational (epistemological reflection and praxeological). We will approach the theories of the autonomy and the independence thus of Wedemeyer (1971), the theory of the incarnate industrialization [2] by Otto Peters (1988) or the theory of the interaction and the communication developed by Holmberg (2003) [3] that constitutes itself from afar a powerful vector of the training. He/it appears therefore important enough that the formation hybridizes permits us to raise the features and the determinants of conceptualization and formalization of the systems of this mixed formation. Our attention is carried more especially on the gotten results that can be reintroduced in the gait of formation engineering in e-learning while respecting the instrumental aspect of the virtual environment.

The second approach is like the steps of research-development. It is about leading the development of the deliverable mediatized in content of formation and activities educational support of teaching jointly. It is therefore about a prospective gait that has the tendency to erase the border between the teaching presently and the training of new expertise on line professionalizing.

3 METHODOLOGY PRECONIZED

Between contingency and choice, between description and reconstruction of a methodological setting of conceptualization of the sessions of formation of the professionals, this descriptive analysis and exploratory appears truly in the reflection of Jean Paul Sartre when it says “the man is characterized above all by the overtaking of a situation, by what it succeeds in making what one made it”.

We note in the analysis of the state of the place, that the pedagogy of the adult learners represents the “poor parent” of the device Blended learning, that is more technical-centric than pedagogical-centric thus. In this perspective he/it appears discriminating to put the focal on the determination of the educational and didactic concepts of references, the devolved place to the adult learner and the modelling of the procedures of screenplay of the objects of trainings and the steps tutorials of accompaniment.

This work of survey and research is in narrow and symbiotic relation with my socio-professional positioning descended of my trajectory of consulting-formative in enterprise and my teacher’s course—formative to the superior normal school of Casablanca the Hassan University 2. These two orientations continue in a third tendency that presents the specificity of my research works in progress in engineering andragogy in the course professionalizing “academic student’s doctoral thesis”.

The present research is a research of development of the devices and convenient of hybrid formation, that can provide to the field of the initial formation the conceptual bases of construction of the mixed formation devices, improving the quality of expertise’s constructed thus. We write down in an analysis deepened of the opportunities offered “benchmarking” of the hybrid models and their impact professionalizing the student’s profession. The specific question of the survey expresses itself of the following manner: facing the difficulties of massification and unhooking of the formation academic presential. What logical of location, construction and formalization of the mixed formation devices also adapted either. They and that encourage the development of the professional expertises of the academic students. What steps of development and conceptualization of the teaching blended specific learning to the context of the faculty of the Ben sciences M sik? To conceive an engineering of the devices of formation of the adult learners in a specific context of Blended formation as support of training, of orientation of the courses of formation and academic output. The development of this point allowed us to put a scheduling rational of the process of production of the deliverable in place material and immaterial on line, the formalization of the procedures

bets in .over in the instrumentation of the numeric platform and the standardization of the norms of the teaching Blended Learning: Variable of omen, product and program is inseparable of the specificity of the training” presently context or from afar.” The report of fact that we could observe, underlined that the courses of formation differentiated translate the needs and the objectives of distinctive formations according to the environment of training. The adult learner appears then like an administrative actor, who elaborates himself his/her/its own course of formation and of which the valorization of his/her/its experience and his/her/its intrinsic incentive bound to needs of masteries and self-improvement. The conception of the modelling of the two environments of trainings would allow us to put the designs of the different units of teachings that encourage the development of the professional expertise’s forward. This modelling of the objects of study drives the learners adulate to mobilize some answers adaptive and systematic of the internal and external resources of every fashion of training, while trying to establish the ties and the bridges of continuity and rupture between these two fashions of trainings. The conceptualization of the hybrid devices makes resort to a consensual approach: “behavior-socio-constructivist and interactive of the training” that the rich and varied environments, centered on the personal professionalization, exploring the global content of the two devices and putting in relief the mechanisms of an intelligent guidance.

The objective of this survey exploratory and descriptive was to understand better and to surround the foundations of the processes of instrumentalisation of the engineering of formation professional of the adult learners in the environment Learning blended. For it, we studied the educational and didactic activities of conception, realization and assessment of the devices presently and from afar, whose crucial stake is the formalization of these devices according to an arsenal of articulate methodological steps adapted to the academic teaching context.

The object of this research aims to articulate way very strong two measurements that are paradox and antinomy: the theoretical dimension [4] and the operative dimension, The crossing and the complementarity of these two aspects contributes to the setting in relief of a conceptual setting constructed of the key concepts of reference, that encourage in a first time the exchanges between the theoretical foundations”. Epistemological” of the process of formation engineering and the pragmatic steps [5] Pastré P. (2004) and operative and that incites in a second time the users to appropriate these instruments and to use them in an adaptive and efficient way.

The central question of research expresses itself of the following manner: facing the difficulties of location of the clear and explicit logics of construction and formalization of the academic courses of formation. What steps of development, formalization and conceptualization of the devices of hybrid trainings?

The setting that served to the analysis of the qualitative and quantitative data of research consists of four main measurements: The engineering of formation and expertise's, the management of the project, the theories of trainings experiential [6] Kolb, D.To.1985 and the approaches andragogical. These measurements are to put directly on the one hand in relation with the theories of interactivity and connectivity of the technological instrument

4 CONCLUSION

The adoption of the technologies of information and communication modified the approach of the formation experiential considerably, notably because of the apparition of the generation Z of the young adults that privileges the fashion par excellence 4C "communication. Connection. Collaboration. Creativeness" as gait of entry to the training of sense, desire and interest. Thus, in light of the data collected the recourse to the devices of hybrid training in the higher education comes to palliate the limits of the teaching presently and to answer the real needs of the adult learners.

We will attach to wonder about the question of research according to which: "What correspondences are necessary here to establish between the effects of the engineering andragogical and development of expertise's professional" academic output, "at adult learners, in situation of training in a context of formation mixes"? Does the relative under-question to research put the accent on the verification of the nature of the reports between the engineering andragogical and do the logics of screenplay of the hybrid teaching "Exist it an engineering andragogy clean to the formation of the adult learners, in situation of training, in a specific context of hybrid formation?" What correspondences to observe between the engineering andragogical and the output of the courses of formation of adult learners, in a specific context of formation Blended Learning? How to characterize the modelling of the objects of studies and activities andragogical and didactical-educational of the process of the professionalization of the courses academic, and this in an environment Learning blended?

The present doctoral research wants to contribute to the advancement of the knowledge in engineering andragogical, while trying to describe and to surround his/her/its theoretical foundations and

praxiological in the mixed training environment, while trying to identify the nature of the relations that becomes knotted between the different variable of the process teaching. Training, as well as the measure of their impacts on the process of professionalization of the students. In the setting of the training in presently, the question of the determination of the aspects of adult's formation has been demonstrated by numerous theoretical and empiric research. He/it appears therefore applicable enough and innovating that the formation Learning blended constitutes a vector of consecration of the values of autonomy and responsibility in the Co-construction of expertise's communicational, methodological and technological in the different situations of trainings computerized "national charter of education and the formation."

Use combined of the teaching presently approaches and enrolls from afar in the gait of continuous improvement of the quality of the trainings school "paradigm of change", indeed the relevance of this Blended formation gait questions the preoccupations and the challenges that the education system Moroccan should raise, notably in relation to the economic and technological internationalization, the adaptation and the interaction with the environment, the innovation and the academic educational renovation, the setting up of new technologies of information and the communication and the good governance and quality of the trainings of the adult learners. It is for it that, we will push from afar on the tools and the techniques of the systems of information and the framing tutorial as well as the engineering andragogy—educational like springboard of formation of the courses professionalizing in the stake to allow the academic system to acquire the determining factors of the essential competitiveness, in occurrence those relative to the technological and human resources to face an omnipresent competition of the internationalization of the market.

He/it appears therefore that the presented theoretical knowledge and research on the engineering andragogical is again in the embryonic phase. Even though several research descriptive and exploratory was interested in the formations to distances like alternative of solution coveted permitting modeler the objects of study and the didactic activities in the computerized environments of training, the results don't seem all in agreement with the foundations praxeological of the paradigms metacognitive and socio constructivists. Certain fashion of training tried to solve the problematic of on line unhooking, massification or democratization of access to the knowledge technical and technological. Other studies focus the attention on the hold in account of the specificities of the learners adult "styles of trainings" in the formations from afar,

in order to increase efficiency of the formation process [7] (Diaz & Cartnal 1999). However the different objects of study show difficulties of consensus, of homogeneity and harmonization of the concepts and methodologies used.

To the term of our doctoral research, we would like to keep to conclude conceptualization and the formalization of the scripts didactics and didaxologie (use and transfer of the scientific data in the gait of teaching). He/it belongs us therefore to justify from the theoretical knowledge presented the arguments fundamental of instrumentalisation of situation of teaching and training of the hybrid teaching in relation with: 1) the training at the adults in narrow report with the principles andragogical of Knowles. 2) The main features of the training in the formation blended structural Learning essentially on the theories of the industrialization and interactivity. 3) The contributions of the professional didactics concerning the courses of professionalization.

The finality aimed of this research survey is therefore at least double. On the one hand, she/it addresses the academic teachers, inventors of the hybrid environments, directly and try to underline the necessity to conceptualize and to formalize the academic courses of formation professionalizing varied. But she/it also addresses the adult learners as gait of incentive, of independence and autonomy of training by report domain the traditional teaching.

The consistency of the hybrid programs offered should be conceived in his/her/its measurements of exchanges through a better quality of interaction and communication, notably in the setting of work intelligent interactive tutorial. He/it takes of it out again therefore that the teaching Learning blended, would be constructed of device of formation privileging a new approach of scheduling where the technology constitutes the elements key of the success of the training of the liberty of choice of the content, of the quality didactic activities, of the educational methods, of the rhythms of work and fashion of varied training (synchronous or asynchronous).

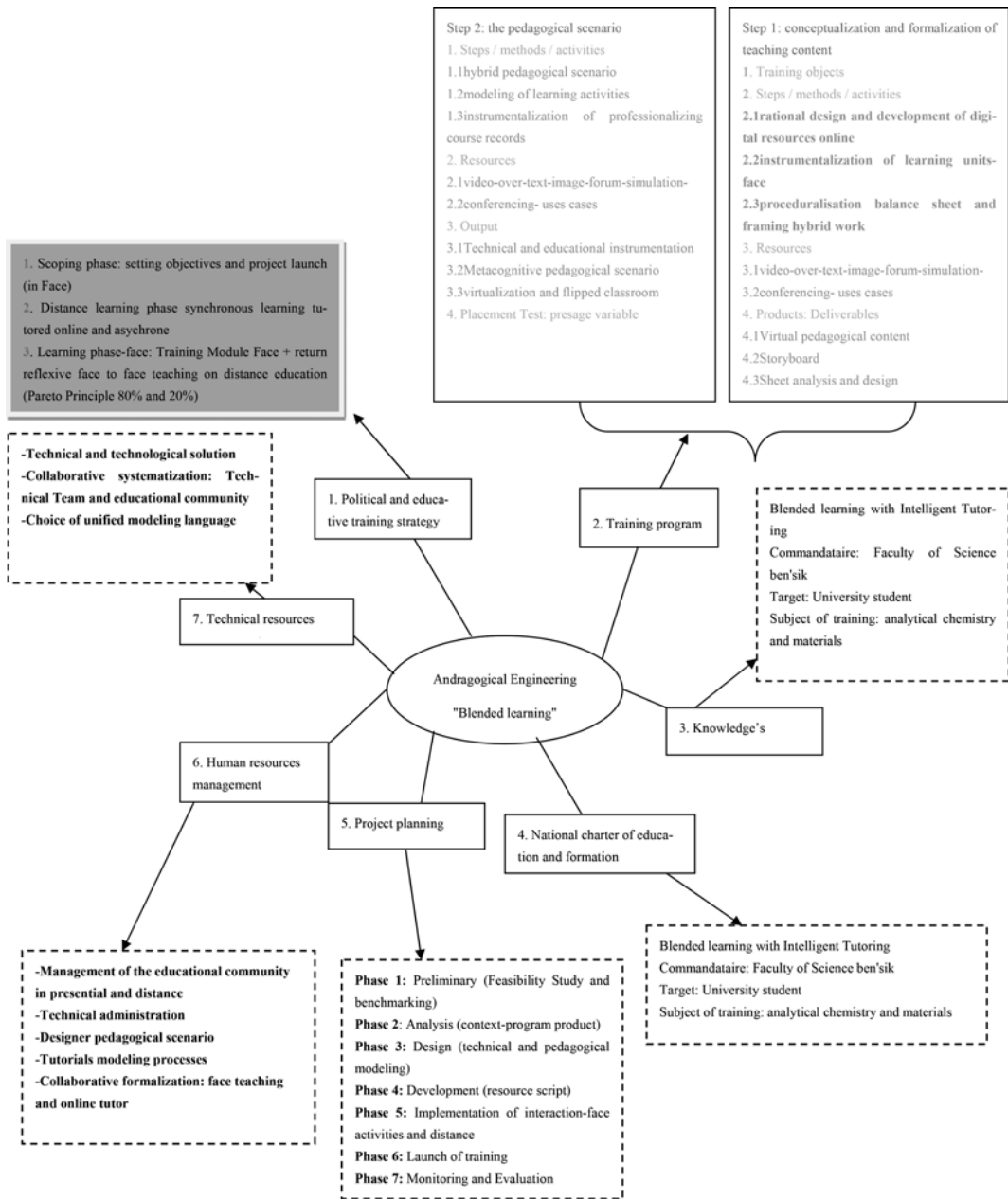
He/it appears therefore alarming enough that the formation from afar and the teaching in presently of the Blended Learning is not completely distinct and is part of a dynamic and interactive educational continuum that goes from an initial situation of training very framed toward a situation of self-culture founds on the principles of autonomy and liberty. We write down here in the process of development of the supple situations of training to the breast of which the accent is put on the exploration, the discovery and the training with his/her/its equals to solve some problems complex and unpublished, in a word of the actions conscientization, reflexive (Depover, Marton, Giardina 1998) and actualized in an environment interactive multimedia.

Our warning was inspired from the subjects of [8] Minder, M. 1999 "he/it is not time to reconcile the currents cognitivists and behaviorist, to break with the fundamentalisms and to place the knowledge to the learner's service".

Admitting the assumption according to which the acquirements will be of as much more steady and efficient that they will be constructed according to the crossing of the different theoretical contributions. In this setting the development of this point places us in a reflection axiological and "didactical-educational" praxiological based on the cohabitation and the complementarity of the teachings (presently and from afar) while using the applicable elements of every fashion of formation in direct relation with the referential of expertise's and the academic educational norms. There is in every theory that founds the hybrid teaching and the engineering andragogical professionalizing the principles and the operative schemes [9] Vergnaud G. (1990). who will be capable to establish the diagnosis of the needs of the adult learners, to formulate some objectives and to identify the material and human resources for the development of the socio-professional expertise's. Of this fact he/it is incontestably necessary to lean on the engineering andragogical to frame and to optimize the steps of teaching, the didactic tools, contextualizing the modes of accompaniment (guidance, facilitation and mediation [10] Jaillot 2004 and of educational regulation of the fashions distance education or presence according to the human and material context (ecological approach) while trying to answer the problematic of the inflation and the obsolescence of the knowledge.

Facing the world of the inflation of the knowledge of the tracks of research future that could prove to be interesting for the domain of the formation engineering of adulates academic, would be to identify if the subjective experience would be one fashion of privileged training in the on line formation context. He/it would be also applicable' after having examined the the modelling of the formation courses identified with the academic output professionalizing.

The academic output passes therefore by a better hold in account of the differences individual experiential. Concerning teaching in Learning blended, he/it would also import to conceive progress of training differentiated according to the needs of generation Z of connection and creativeness. This instrumental modelling of the virtual environment but also of the presently <Face to Face> according to the law of the Pareto 80. 20 to give to the psycho-emotional dimension as vector of development of an emotional and transcendent intelligence.



Grille 1. Instrumentalisation and modelling of the engineering andragogique in devices blended learning.

REFERENCES

[1] Merriam, S.B. and Caffarella, R. (1991). Learning in adulthood. To understanding guide. San Francisco, THAT: Jossey-Bass. Golden, S. and Basque, J. (1998). The computerized training environment concept. Reviewed from afar of the education, 13 (1), 40–56.

[2] Wedemeyer C.A., (1971) Independent study. In Lee C. Deighton (Editor Chief in), the Encyclopedia of Education, flight 4 (p. 550). New York: The MacMillan Co12.

[3] Holmberg, B. (2003). To theory of education based outdistances one empathy. In M. G. Moore, W. G. Anderson, Handbook of outdistances education.

- Mahwah, N-J: Lawrence Erlbaum Associates, 79–86.
- [4] Peter OH. (1998). Learning and teaching in outdistances education Analysis and interpretation from year international perspective. London: Kogan Page 13.
- [5] Pastré P. (2004). “The role of the pragmatic concepts in the management of the situations problems: the case of the regulators in plasturgie”. In R.15.
- [6] KOLB, D.To. 1985. The learning, style inventory: self, corin inventory and interpretation booklet. Boston: Mass, McBer Company and (1st éd. 1981).
- [7] Diaz, D (1999) and Cartnal Comparing Student Learning Styles in year online Outdistances Learning Class and year equivalent Class one-campus. *College Teaching* 47 (4), pp. 130–135.
- [8] Minder, M. 1999. Functional didactics. Brussels: Of Boeck University.
- [9] Vergnaud G. (1990). “The theory of the conceptual fields”. *Research in math didactics, flight*. 10, n° 2–3, p. 133–170.
- [10] Jaillet, has. 2004. The school to the numeric era. Paris: the Harmattan.

A general scheme for MRI images' segmentation based on slice-by-slice learning

Ismael Kone & Lahsen Boulmane

Faculté des Sciences—Université Moulay Ismail, Meknès, Morocco

Mohamed Radouani & Benaissa Elfahime

ENSAM—Université Moulay Ismail, Meknès, Morocco

ABSTRACT: Machine learning models experience a great success in MRI images' segmentation. Although training these models comes at a heavy cost because an important amount of data must be labeled manually. Therefore lowering this cost is more than welcome. We propose a general scheme algorithm for binary MRI images' segmentation to nearly drop out this expensive task. Only the first slice needs to be manually labeled to train the system which then infers the second slice's segmentation. Inference result updates the system's knowledge which then infers the third slice and so on till the last slice. As our study is still in process, we tested this approach with a mixture of Gaussians for a femoral bone's segmentation. Despite its limits, results are very promising. Also a recent work followed the proposed scheme using a more elaborate model without formalizing it as we do. Their results demonstrated its effectiveness for the placenta's segmentation. Systems that implement this scheme have a high potential to be adaptive to different segmentation tasks because they restart learning on every new task.

1 INTRODUCTION

Numerous Machine Learning (ML) models addressed the problem of anatomy segmentation in MRI images and yield good results. However these approaches have two shortcomings:

- The need of huge amount of data manually labeled for training the ML model before deploying it for use.
- The lack of flexibility because each segmentation requires specific model to be trained.

To illustrate the first shortcoming, consider the liver segmentation. The training step may require MRI images from 10 subjects. MRI images from a subject produces 10 slices that have to be annotated manually by experts. Finally $10 \times 10 = 100$ slices must be annotated manually and that represents obviously too much work and time.

For the second problem, we keep the previous example and we suppose the ML model is already trained to segment the liver. But after a while, we are in need of segmenting a femoral bone. From the best of our knowledge we can't extend the previous model for this task. Thus we must train a new model with the same burden of training data mentioned.

In this paper we propose a flexible scheme for segmenting any organ in MRI images. Our contributions are:

- A description of the segmentation problem in the context of MRI images based on simple human observations. The proposed scheme is a direct consequence of this description (section 2).
- A simple implementation of this scheme for femoral bone segmentation to show this method is promising (section 3–4).

For the last contribution, we didn't provide enough examples as work is still in process. But we will discuss this approach and relate it to other works (section 5).

2 DESCRIPTION AND FORMULATION OF THE PROBLEM

2.1 Description

Our objective is to segment a specific organ from a body part MRI images. MRI in a clinical context produces a stack of 2D images called slices of the given 3D volumetric part of the body. These slices can be seen as a discretization of the body part along. The distance between two consecutive slices is called inter-gap distance which is usually less than 3 mm. Figure 1 shows some MRI slices.

At first glance, we notice smooth changes of the slice's contents from one to another, especially for two consecutive slices. The short inter-gap distance explains these smooth changes. When we observe

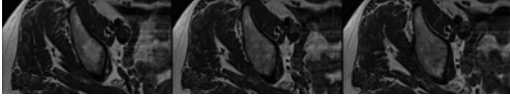


Figure 1. Consecutive slices from femoral bone MRI.

the first slice and then we look at the second, we directly make associations between both slices. Specifically, we may say for instance, a given area in the first slice is a bone and another area in the second slice has approximately the same appearance and position than previous slice’s bone. Therefore this area in the second slice is the same bone. And we do it through all slices so that we can identify the same bone in all of them despite of changes on the shape or aspect.

Implicitly, we learn the bone from the first slice. Then, as we observe the second slice we infer areas that are likely to be next part of the bone previously seen. This is done based on visible features: bone’s position in the slice, its shape and gray intensities distribution over its surface which may present textures. At this step we refine or enrich our knowledge of the bone appearance and we observe the next slice and infer again in the same way. These observations guided us to emulate the same process by designing a machine learning approach including uncertainty assessment to drive the inference.

2.2 Formulation

We model the problem as a classification where each pixel of a slice will be labeled as the desired organ or not. Hence we restrict ourselves to the binary case nonetheless extension to multi-class classification is straightforward. A formal presentation of what we described is summarized in the algorithm in Table 1. The input are MRI slices and the first slice manually segmented i.e. the area of the slice that is the desired organ namely O_1 . We start by learn from it and the background B_1 and that knowledge is named K . From that point we loop over next slices where organs are defined by a set of pixels that satisfy a criterion: the probability of belonging to the organ of the current slice with the current level of knowledge K must be greater than α . This parameter allows us enforce a level of certainty on the system and must be at least 0.5. It implies the probability of belonging to the background is less than 0.5. Normally, next step should be post-processing to suppress outliers and add inliers that are respectively background pixels misclassified as bone and vice-versa due to inference errors. But focus must be put on building a robust model that reduces the number of outliers and inliers to nearly zero. Then we refine or enrich

Table 1. General scheme algorithm for MRI images organ segmentation.

Algorithm: Input $(\{S_n/n=1, \dots, N\}, O_1, \alpha \geq 0.5)$

1. $K \leftarrow$ Learn O_1 and $B_1 = S_1 \setminus O_1$
 2. Loop over $n = 1, \dots, N - 1$:
 3. $O_{n+1} \leftarrow \{\text{pixel} \in S_{n+1} / \text{Probability}(\text{pixel} \in O_{n+1} | K) \geq \alpha\}$
 4. Refine K with O_{n+1} and $B_{n+1} = S_{n+1} \setminus O_{n+1}$
 5. Output: $\{O_n/n=1, \dots, N\}$
-

the knowledge K with new data coming from lastly segmented organ and background. At the end we output organs’ set.

Now main questions are: How can we implement this algorithm concretely? How learning is done? What are really these probabilities and how are they computed?

3 METHOD

Before addressing the learning and inference, we must make a choice of the classification model to use whether generative or discriminative. In fact subsequent learning and inference algorithms depend completely on the model (Prince 2012).

3.1 Model

Experiences should guide our choice but as we are in early work, we decide to use a generative model because of two interesting properties. One is its ability to interpolate missing information in training or test data and the other is its relatively easy prior knowledge incorporation in contrary to discriminative models (Prince 2012). We decide to use the Gaussian Model Mixture (GMM) because it can model multimodal data densities and has been well studied.

3.2 Learning and inference

In our setting, we are treating binary classification so we use two GMMs: one for the organ and the other for the background. Training data are in pairwise $\{(x_i, y_i), i = 1, \dots, N\}$ where x_i represents the feature vector of a pixel and y_i represents its label whether 0 or 1 indicating respectively that the pixel is in the background or in the organ. Hence, we used data where $y_i = 1$ to learn the organ and data where $y_i = 0$ for the background. In other terms the organ’s GMM is modeled by the likelihood $\text{Probability}(x | y = 1)$ and the background’s GMM is modeled by $\text{Probability}(x | y = 0)$ where x is the vector feature of a pixel. Expectation-Maximization (EM) (McLachlan & Krishnan 2008)

algorithm is applied to learn parameters of each GMM from data.

To infer a new datum x , we use Bayes' rule to compute posterior $Probability(y = 1 | x)$ which corresponds to the probability stated in the general scheme algorithm (Table 1).

If this posterior is superior to α then x is labeled to the organ class, else it is a background pixel.

3.3 Features

According to the description (section 2), elements that allow us to infer are: the organ's position, its shape and gray intensities distribution over its surface. To remain simple, we decide to use only the pixel coordinates and gray intensity. However all possible features are derived from both.

4 EXPERIMENTS AND RESULTS

We used a laptop Intel i3 CPU core on Ubuntu 12.04 environment with python language programming. We used also OpenCV (Bradsky 2000) python binding for image manipulation and GMM implementation from Sklearn library (Pedregosa et al. 2011).

Test images are MRI images from human femoral bone (Fig. 2).

We used a simple routine for segmenting first slice based on dominant object detection algorithm (Kone et al. 2016), used $\alpha = 0.5$ and 15 components for each GMM (bone and background).

Results are quite remarkable (Fig. 3) despite the low level of features used. In this experiment we

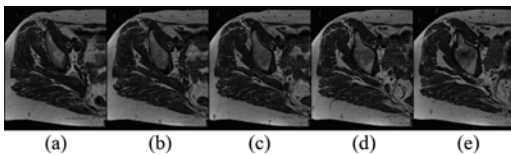


Figure 2. Test images: sample of slices from femoral bone MRI.

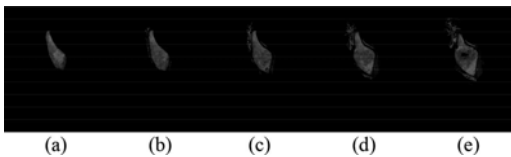


Figure 3. Results after applying the general scheme algorithm on test images (Fig. 2) using GMM with 15 components as model. Segmentation uses only the previous slice knowledge.

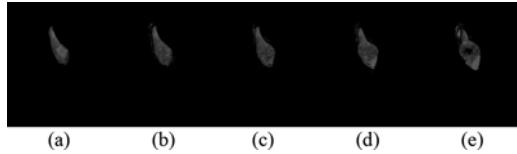


Figure 4. Results after applying the general scheme algorithm on test images (Fig. 2) using GMM with 15 components as model. Segmentation results based on first slice learning only without changes knowledge acquired.

didn't enrich the knowledge K . Instead we replace K by current knowledge so that only knowledge from the previous slice is used to infer the current one's segmentation. This is a limit of the EM algorithm which allows only batch learning. Outliers and inliers add up as we move through slices. Figure 4 presents results when we learn only the first slice without updating knowledge for remaining slices' segmentation. We notice a few outliers because the slice learned has no outliers. That points the need to build more robust model to have less noisy segmentation result. Besides, outliers and inliers handling is application specific thus prevents the system to be adaptive.

5 RELATED WORK

Active learning (Burr 2009) studies the problem of using the smallest manually labeled data as possible to train a robust system. As we stated, manual labeling is expensive so this problem has been around for a while and is still ongoing. We stress it extremely in the context of MRI images segmentation which promotes it. Our problem formulation aims to reduce the set of manual labeling to one. Some works adopted almost a same approach as ours. Lorigo et al. (1998) proposed a method for bone segmentation in knee MRI using active contours. In the process, the same initial contour was used for all slices except first and last slices because there were few changes. That's exactly what we did in Figure 4. More recently, Wang et al. (2015) developed a system for placenta segmentation using a slice-by-slice approach namely Slic-Seg as we devised our general scheme algorithm. They used an online random forests model to learn on the fly. Results show robustness of the method compare to methods like Geodesic framework (Bai & Sapiro 2008), generalized gradient vector flow (Chenyang & Prince 1998) and GraphCut (Boykov & Jolly 2001) and less variability between users. This is a clear pointer that following such a scheme is really promising in MRI images segmentation.

6 CONCLUSION AND FUTURE WORK

We addressed the cost of manual data labeling in machine learning method for MRI images segmentation and presented a general algorithm scheme to face this problem. Not only following this scheme can nearly drops out this cost but also could allow the system to be more flexible. Some work show its liability in particular tasks. We project to pursue this study by addressing the problem of building a robust model to get a high accuracy rate. Then we will study its adaptivity by experimenting different organ segmentations in MRI images.

REFERENCES

- Gary, Bradski 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.
- Geoffrey, J. McLachlan & Thriyambakam, Krishnan 2008. *The EM algorithm and extensions 2nd edition*. Wiley.
- Guotai, Wang & Maria, A. Zuluaga & Rosalind, Pratt & Michael, Aertsen & Anna, L. David & Jan, Deprest & Tom, Vercauteren & and Sebastien Ourseli: Slice-by-Slice Segmentation Propagation of the Placenta in Fetal MRI Using One-Plane Scribbles N. Navab et al. (Eds.): MICCAI 2015, Part III, LNCS 9351: 29–37.
- Ismael, Kone & Lahsen, Boulmane & Mohamed, Radouani & Benaissa, Elfahime 2016. Automatic femoral bone segmentation in Hip MRI images. *Advances in Information Technology: Theory and Application* (2489–1703) 1(1): 75–78.
- Liana, Lorigo M. & Olivier, Faugeras & Grimson, W. E. L. & Renaud, Keriven & Ron, Kikinis 1998. Segmentation of Bone in Clinical Knee MRI using Texture-Based Geodesic Active Contours. *Medical Image Computing and Computer-Assisted Intervention—MICCAI' 98*, Lecture Notes in Computer Science (0302–9743) 1496: 1195–1204.
- Pedregosa *et al.* 2011. Scikit-learn: Machine Learning in Python, *JMLR* 12: 2825–2830.
- Settles, Burr 2009. Active learning literature survey. *Computer Sciences Technical Report* 1648, Madison: University of Wisconsin.
- Simon, J. D. Prince 2012. *Computer Vision: models, learning and inference*. London: Cambridge University Press.
- Xu, Chenyang & Jerry, L. Prince 1998. Snakes, Shapes and Gradient Vector Flow. *IEEE TIP* 7(3): 359–369.
- Xue, Bai & Guillermo, Sapiro 2008. A Geodesic Framework for Fast Interactive Image and Video Segmentation and Matting. *IJCV* 82(2), 113–132.
- Yuri, Y. Boykov & Marie-Pierre, Jolly 2001. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. *ICCV* 1: 105–112.

FreebaseViz: Interactive exploration of freebase schema using query-driven visualisation

Mahmoud Elbattah

College of Engineering and Informatics, National University of Ireland, Ireland

Mohamed Roushdy, Mostafa Aref & Abdel-Badeeh M. Salem

Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt

ABSTRACT: Visualization is attaining a growing recognition as a pivotal part of the data analysis process. Visualization-based solutions are increasingly used to adequately explore and communicate understanding of large-scale datasets. This paper presents a web-based visualisation tool, named FreebaseViz, for visually exploring the schema of Freebase. The visualisation design is built upon node-link network layouts, which can facilitate exploring connectivity, visual search and analysis, and visualising patterns underlying the schema graph. FreebaseViz allows users to interact with the schema visualisations to filter and drill into lower levels of detail, and highlight subsets of the schema graph. Furthermore, a graph database-oriented approach is embraced in an endeavour to boost the visualisation query-ability using graph-based query operations. A set of visualisation scenarios were constructed in order to demonstrate the applicability and usefulness of the tool. The paper proceeds to draw observations based on the schema visualisations, which attempt to provide a deeper understanding of the Freebase schema model. The visualisations mainly infer that there are a few super-connected nodes within the schema graph, which can be interpreted as that the Freebase schema resembles the structure of a scale-free network.

Keywords: data science, schema visualisation, user interfaces, freebase, graph database

1 INTRODUCTION

Visualisation was ideally described as the transformation of the symbolic into the geometric [5]. Various benefits can be attributed to data visualisation. In contrast to text-based means, the interpretation of visual formats happens immediately in a pre-attentive manner. Further, the pictorial representation of data can help answer or discover questions. The usefulness of exploratory data analysis using visual techniques was early introduced in John Tukey's landmark textbook *Exploratory Data Analysis* [1].

However, visualisation has gained a particular significance in the wake of Big Data. For instance, the Linked Open Data (LOD) [15] initiatives enabled a vast amount of data published in freely accessible datasets, such as Freebase [2] and DBpedia [3]. Exploratory visualisations have become an imperative in order to discover and summarize the main characteristics of such large-scale datasets. In this context, the paper addressed schema exploration using interactive visualisation with a particular focus on Freebase.

Freebase includes a great diversity of structured data imported from different sources such as Wikipedia, MusicBrainz and WordNet [14]. The breadth and complexity of the Freebase schema were acknowledged by other studies. For instance, it was recognised [18] that even expert users can need to perform a lot of effort to understand the Freebase schema and find the relevant parts to be able to write well-structured queries. While the study [19] emphasised that the complexity of Freebase schema makes it difficult to query and explore.

In light of that, the study introduced a tool that can support an exploratory visualisation of Freebase schema, named as "FreebaseViz" [23]. The visualisation was provided with interactivity and query-driven capabilities. Concretely, we attempted to make contributions in two aspects. First, the developed tool is claimed to provide a better understanding of the Freebase Schema and how data is organised within its massive repository. Second, the study inspected the potential usefulness of graph databases with regard to query-driven visualisations in terms of enabling graph-based queries.

Although the paper only endorsed the schema of Freebase, it can merely be considered as an exemplar of large-scale schemas, and the tool can be applicable or extensible for similar use cases.

2 RELATED WORK

This section overviews exemplar visualisation tools with reference to two aspects. First, the tools that utilised interactive visualisation methods in order to explore database schem as in general. Second, the tools that attempted to visually explore Freebase in particular. Table 1 summarises the surveyed visualisation tools.

Based on the reviewed tools, a set of considerations can be highlighted to the best knowledge of the authors. Apart from a few endeavours such as Thinkbase [11][12], the literature laid a little emphasis on studies that used visualisation methods for the purpose of understanding Freebase data. Furthermore, we could not find a similar study focusing on exploring or

understanding the Freebase schema from a visualisation perspective. On the other hand, the literature obviously lacked attempts to employ graph databases in the context of query-driven visualization.

3 CONCEPTUALISATION OF FREEBASE SCHEMA

Generally, the Freebase schema is expressed in terms of “Types” and related “Properties” [20]. The Types are grouped together to form “Domains”. Similarly, the Domains are grouped together to form broader “Categories”. Table 2 explains with examples the breakdown of the schema. In view of that, the Freebase schema model can generally be conceived as inter-linked graphs, which collectively comprise the entire schema graph. Figure 1 illustrates an example describing the “Politician” Type and its relationships in the schema graph. The study utilised the schema graph previously extracted from Freebase by [17].

Table 1. Related visualisation tools.

Tool	Description
NakeDB [9]	The NakeDB tool endorsed the dynamic interaction of database schemas using visualisation techniques. NakeDB was a Java tool that used the Prefuse toolkit [10]. The database schema had to be parsed into XML files to be processed by Prefuse.
Schemr [13]	Schemr was built as a search engine to search and visualize database schemas in a metadata repository. The search results were visualised in an interactive web application, allowing users to visually explore a schema. The Schemr’s GUI supported interactions, including panning, zooming, and drilling-in.
Graph Charter [16]	Graph Charter proposed a method that combined graph browsing with querying to improve the capacity of visual inspection. The visualisation of Freebase knowledge graphs was used as a case study.
Thinkbase [11][12]	Thinkbase is considered as one of the most significant endeavours for developing an exploratory visualisation tool of Freebase data. Thinkbase could extract the contents and semantic relationships from Freebase and visualises them using an interactive visual representation. However, Thinkbase did not endorse the schema of Freebase.

Table 2. The components of the schema model of Freebase.

Element	Description	Example
Type (n = 2,120)	Denotes an Is-A relationship about a topic.	The “Albert Einstein” topic is a type of Person.
Property (n = 6,851)	Define a Has-A relationship between the topic and the value of the property.	Date of Birth, Place of Birth, and Places lived for topics of typed as Person.
Domain (n = 82)	A collection of Types that share a namespace. There are 82 Freebase-defined Domains.	Education, Music and Religion.
Category (n = 9)	A grouping of related Domains, including 9 categories.	Science & Technology, Arts & Entertainment, and Society.

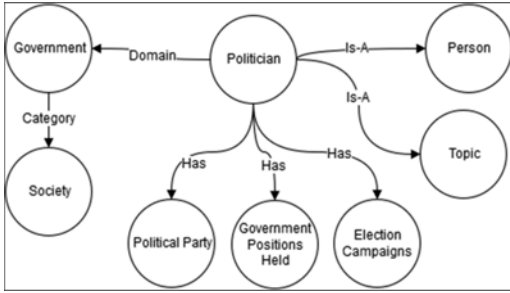


Figure 1. Example of the Freebase schema structure: The Politician Type. The example shows how the Types, Properties, Domains and Categories are organised within the Freebase schema.

4 FREEBASEVIZ ARCHITECTURE

The FreebaseViz tool was developed as a web-based application on top of the .NET framework 4.0, and Neo4j graph database. The architecture of FreebaseViz is sketched in Figure 2. Specifically, the architecture was organised into four layers as follows: i) Storage Layer, ii) API Layer, iii) Interactivity Layer, and iv) Visualisation Layer. The first layer served the purposes of storing and querying the Freebase schema, whereas the schema was initially persisted as a single graph into the Neo4j graph database. The second layer endorsed the API functionality, where data queries were handled using REST-ful web services. The queries were coded using Cypher, the Neo4j query language, and results were returned in JSON format. The API layer utilised the open-source Neo4j Clientlibrary [21]. The third layer provided the user interface and interactivity features through an ASP.NET web application. The top layer produced schema visualisations, which were rendered using the Javascript library of VivaGraphJS [4].

The FreebaseViz tool was designed with the aim of providing Query-Driven Visualization and analysis (QDV) of the Freebase schema. The term QDV was coined by [6, 7] to describe the combination of high performance query methods with visual data exploration methods. Towards that aim, the study adopted a graph database-oriented approach for storing and querying the schema graph. The graph database played a principal role for enabling graph-driven query operations. Particularly, using a graph database facilitated building queries based on graph patterns with different levels of complexity. For instance, extracting a sub-graph from the schema graph can be seamlessly implemented using the query capabilities of graph databases.

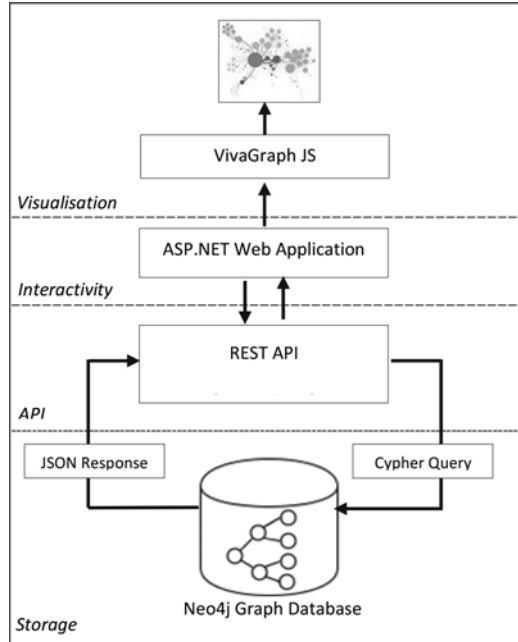


Figure 2. Overview of the FreebaseViz architecture.

5 VISUAL DESIGN

5.1 Layout view

The user interface of FreebaseViz featured two panels as highlighted in Figure 3. On one hand, the left-sided panel served two functions including: i) Displaying graph summary information, such as the number of nodes and links, and ii) Providing interactivity functions, such as zooming and filtering for example. On the other hand, the right-sided panel provided a workspace for users to explore schema visualizations of Freebase. The schema visualizations were constructed as node-link diagrams. Specifically, the graph nodes depicted the Freebase Types, while edges represented the Is-A relationships. The node-link layout provided an intuitive way to explicitly visualize the schema structure and underlying relationships. Moreover, FreebaseViz utilised a force-directed layout in order to reduce the visual clutter that can be caused while visualizing a large number of nodes and edges.

5.2 Visual encoding

A. Nodes

The graph nodes represented Freebase Types, such as “Film”, “Book” or “Politician” for example. In addition, every node possessed a set of attributes. These attributes characterised the following: i) Type’s name ii) Type’s domain, iii) Domain

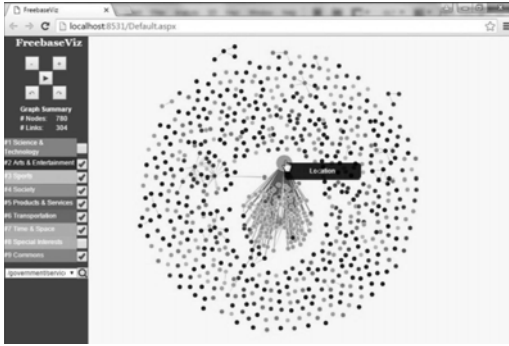


Figure 3. The layout view of the FreebaseViz tool. The left-sided panel contains various interactivity features in addition to displaying graph summary. While the right-sided panel displays visualisations.

category, and iv) Instance count, which represents the number of topics belonging to that Type.

B. Node colouring

The schema of Freebase is basically structured into 9 categories including: 1) Science and Technology, 2) Arts and Entertainment, 3) Sports, 4) Society, 5) Products and Services, 6) Transportation, 7) Time and Space, 8) Special Interests, and 9) Commons. Based on those categories, the nodes were distinctively colour-encoded to visually associate nodes with Freebase categories.

C. Node size

The size of a node was used to visually illustrate an important attribute of the Freebase Types, which is the instance count. The instance count of a Type indicates the number of physical or abstract topics instantiated from that Type. The higher is the instance count, the larger is the node size.

D. Edges

The Freebase schema is centralised around the Is-A relationships as explained in Section 3. The Is-A relationships were described via directed edges connecting the nodes in schema visualisations. For example, the edge linking the nodes of “US State” and “Location” can infer that the two Types are connected with an Is-A relationship.

E. Edges colouring

The schema of Freebase was considered as a directed graph, where the Is-A relationships represented the directed edges. In accordance with the Is-A links, directed edges were coloured with a linear gradient ranging from yellow to red. Specifically, the yellow and red colours defined the source and destination nodes respectively.

6 INTERACTION TECHNIQUES

A visualisation can raise a set of useful questions, however if the visualisation is static it can be hard to answer those questions. Regarding FreebaseViz, the interactivity was indispensable to investigate schema visualisations from multiple perspectives. FreebaseViz implemented two categories of interaction techniques for the purposes of navigation and exploration of the schema graph. The interactivity features were developed in light of Shneiderman’s visual information mantra as “overview first, zoom and filter, then details-on-demand” [22]. This section describes the interactivity capabilities provided by the tool.

6.1 Navigation: Panning and zooming

FreebaseViz utilised panning and zooming for navigating around the visualization space. On one hand, the panning can be performed by dragging the background of the visualization canvas with the mouse left button down. On the other hand, the visualisation can be zoomed easily using the mouse scroll, or by clicking the zoom in/out buttons located in the left-sided panel.

6.2 Navigation: Selection and dragging

The nodes of the schema graph are selectable and draggable. Further, the graph can be automatically re-positioned with respect to a user-selected node by clicking the Play button at the left-sided panel.

6.3 Navigation: Rotation

FreebaseViz provided a 2-D rotation capability of the visualization space. The rotation can help users investigate the schema graph from different angles. The user can control the direction of rotation using two buttons at the left-sided panel.

6.4 Exploration: Tooltips

The nodes in the schema graph had a variety of attributes that can be displayed to the user, whereas FreebaseViz provided tooltips for exploring those attributes. Specifically, two categories of tooltips were designed including: i) On-hover tooltip, and ii) On-click tooltip. The on-hover tooltips displayed names corresponding to the Freebase Types, while cursor moving over graph nodes. Additionally, the on-click tooltip were used to open a more detailed view of the attributes including: i) Instance count, ii) Type’s domain, and iii) Category of the domain, and iv) Type’s properties.

6.5 Exploration: Highlighting neighbour nodes

Finding the neighbours of a node is commonly required in the context of visualisation in order to highlight links or relationships. Further, identifying the neighbouring nodes can considerably assist in understanding the structure and conceptual connections within the schema. In FreebaseViz, this was simply achieved by moving the cursor over a node, and then the node's connections are automatically highlighted with the red colour.

6.6 Exploration: Filter and search

FreebaseViz implemented filtering and search features to enable users to drill down into the schema graph and focus on a region of interest. Two methods of dynamic filtering were provided as follows: i) Filtering by category, and ii) Finding a node (Freebase Type) and its connected nodes. Filtering by category enabled the user to dynamically filter the visualisation with respect to the 9 Freebase categories mentioned in Section 3. The category-based filters can be applied using checkboxes placed at the left-sided panel. In addition, a specific Type and its connections can be visualised by selecting a specific Type from a dropdown list.

7 VISUALISATION SCENARIOS

This section presents three scenarios of schema graph visualisations. On one hand, the interactivity features of FreebaseViz were inspected through the visualisation scenarios. On the other hand, the produced visualisations helped to draw observations on the schema structure of Freebase in particular.

7.1 First visualisation scenario: Finding dominant types

This scenario aimed at identifying the densely connected nodes in the schema graph. The visualisation endorsed the following questions: 1) Are there particular Types that dominate the Is-A relationships in the schema graph?, and 2) If yes, do they all belong to the same domain or category? Figure 4 shows the output of the visualisation. Further findings are discussed in the next section.

7.2 Second visualisation scenario: Category-filtered schema graph

The second scenario utilised one of the interactivity features of FreebaseViz described at Section 6.6, which enabled to filter the schema graph with respect to particular categories. The categories of "Science and Technology" and "Society" were

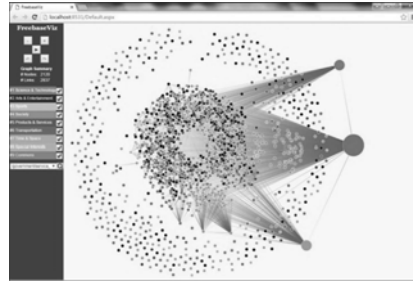


Figure 4. Visualisation Scenario: Finding dominant schema Types. The visualisation reveals that the schema graph is dominated by a few Types represented as the large-size nodes at the far right of the graph. The dominant Types are Person (Green), Topic (Red), and Location (Orange).

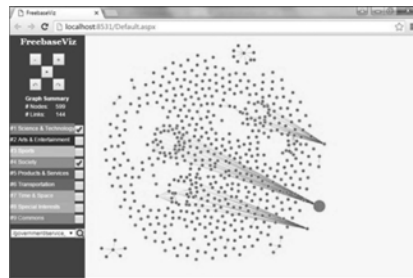


Figure 5. Visualisation Scenario: Category-filtered schema graph. First, the schema graph was filtered with respect to the selected categories on the left-sided panel, which are "Science and Technology" and "Society". The visualisation can help identify the entities that have mutual connection related to the both categories. Further, it can be observed that the "Science and Technology" category (blue nodes) tends to have smaller well-connected communities.

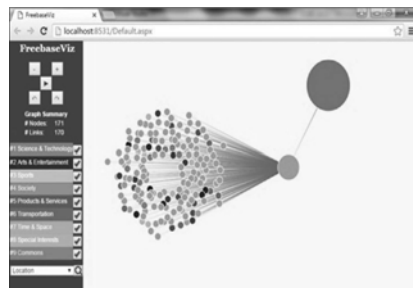


Figure 6. Visualisation Scenario: Type-filtered schema graph. First, the schema graph was filtered with respect to the Location Type, selected from the dropdown box on the left-sided panel. The big orange-coloured describes Location, while the bigger red-coloured describes Topic. The visualisation shows the various categories of Types in relation with Location. However, the "Time and Space" category (orange nodes) tends to dominate the schema relationships with Location.

used as an example for this scenario. The visualisation endorsed the following question: What is the magnitude of connections between the Types belonging to “Science and Technology”, and those belonging to “Society”? Figure 5 shows the produced visualisation.

7.3 Third visualisation scenario: Type-filtered schema graph

The final scenario drilled down into the Type-level relationships in the schema graph. It was aimed to filter the schema graph with respect to a particular Type. The “Location” Type was considered as an example for that purpose. In particular, the visualisation endorsed the following question: Which Types are connected to the Location Type?, and ii) Is there a prevalent category of Types in relation with Location? Figure 6 shows the visualisation.

8 OBSERVATIONS

A set of observations could be drawn from the produced visualisations. In terms of schema structure, the visualisation showed that the Freebase schema structure resembled that of a scale-free network [8], whereas its degree distribution followed a power law distribution. Specifically, a few “super-connected” nodes obviously dominated the Is-A relationships underlying the schema graph, which were Topic, Location and Person. Figure 7 emphasizes the schema structure by plotting the in-degree distribution of the schema Types.

Furthermore, the visualisation showed that the schema graph tended to have densely connected nodes, which can qualify to form communities. However, a considerable proportion of the schema Types had no connections in the schema graph. Those disconnected nodes can be seen as a ring

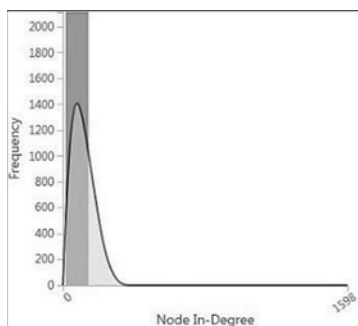


Figure 7. In-degree distributions in the schema graph. It can be observed that the in-degree distribution follows a power-law, which conforms with the visualisation output of the first scenario in particular.

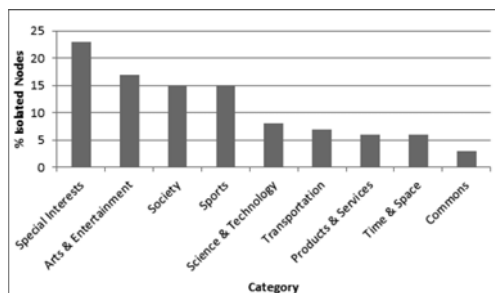


Figure 8. The percentage of isolated Types within the 9 categories of Freebase, which have no connections with other Types. It can be noticed that the “Special Interests” and “Commons” categories have the lowest and highest proportions of isolated Types, respectively.

around the graph, as shown in Figure 4 produced by scenario 1. Particularly, the number of disconnected nodes represented about 22% of the overall schema entities. Figure 8 shows precisely the percentage of isolated nodes within the 9 categories of Freebase schema.

9 CONCLUSIONS

Visualisation presents as an appropriate approach for realising a rich understanding of schemas, especially in the case of large-scale schemas. Insights and context can be drawn by visually exploring relationships, and discovering underlying patterns. In this manner, the FreebaseViz tool demonstrated an effective approach for the visual exploration of Freebase schema. On one hand, the tool provided various interactivity techniques for schema navigation, such as panning and zooming, selection and dragging, and rotation. On the other hand, dynamic filtering and search operations were provided to enable a focused inspection of the schema.

A set of visualisation scenarios were developed in an attempt to discover the schema characteristics of Freebase in a visual manner. The visualisations revealed that the schema resembled the structure of a scale-free network, whereas its degree distribution followed a power law distribution, and a few super-connected nodes dominated the schema graph connections. In contrast, a considerable proportion of the schema Types seemed isolated with no connections in the schema graph.

Furthermore, the study investigated the potential usefulness of graph databases for visualization environments in particular. In this respect, FreebaseViz took advantage of the graph database capabilities in order to realise graph-oriented query operations. Specifically, it can be concluded

that the usage of graph databases can carry special benefits for visualisation utilities, where traversing large-scale graphs or extracting sub-graphs, for example, can be efficiently implemented.

REFERENCES

- [1] Tukey, John W. "Exploratory data analysis." (1977): 2–3.
- [2] <http://www.freebase.com/>
- [3] <http://wiki.dbpedia.org/>
- [4] <https://github.com/anvaka/VivaGraphJS>
- [5] McCormick, Bruce Howard, Thomas A. DeFanti, and Maxine D. Brown. "Visualization in scientific computing." *IEEE Computer Graphics and Applications* 7, no. 10 (1987): 69–69.
- [6] Kurt Stockinger, John Shall, E. Wes Bethel, and Kesheng Wu. DEX: Increasing the Capability of Scientific Data Analysis Pipelines by Using Efficient Bitmap Indices to Accelerate Scientific Visualization. In *Proceedings of Scientific and Statistical Database Management Conference (SSDBM)*, pages 35–44, Santa Barbara, CA, USA, June 2005. LBNL-57203.
- [7] Kurt Stockinger, John Shall, Kesheng Wu, and E. Wes Bethel. Query-Driven Visualization of Large Data Sets. In *Proceedings of IEEE Visualization 2005*, pages 167–174. IEEE Computer Society Press, October 2005. LBNL-57511.
- [8] Barabási, Albert-László, and Réka Albert. "Emergence of scaling in random networks." *Science* 286, no. 5439 (1999): 509–512.
- [9] Cortes-Pena, Luis Miguel, Yi Han, Neil Pradhan, and Romain Rigaux. "NakeDB: Database schema visualization." *Proc. of the APRIL 2008* (2008).
- [10] J. Heer, S. Card, and J. Landay, "prefuse: a toolkit for interactive information visualization," *Conference on Human Factors in Computing Systems*, pp. 421–430, 2005.
- [11] Hirsch, Christian, John C. Grundy, John G. Hosking. "Thinkbase: A Visual Semantic Wiki." *International Semantic Web Conference*, 2008.
- [12] Hirsch, Christian, John Hosking, John Grundy. "Interactive visualization tools for exploring the semantic graph of large knowledge spaces", *Workshop on Visual Interfaces to the Social and the Semantic Web (VISSW2009)*, vol. 443, 2009.
- [13] Chen, Kuang, Akshay Kannan, Jayant Madhavan, and Alon Halevy. "Exploring schema repositories with schemr." *ACM SIGMOD Record* 40, no. 1 (2011): 11–16.
- [14] http://wiki.freebase.com/wiki/Data_sources
- [15] Bizer, Christian, Tom Heath, and Tim Berners-Lee. "Linked data—the story so far." *Semantic Services, Interoperability and Web Applications: Emerging Concepts* (2009): 205–227.
- [16] Tu, Ying, and Han-Wei Shen. "Graph Charter: Combining browsing with query to explore large semantic graphs." In *Visualization Symposium (PacificVis)*, 2013 IEEE Pacific, pp. 49–56. IEEE, 2013.
- [17] Elbattah, Mahmoud, Mohamed Roshdy, Mostafa Aref, and Abdel-Badeh Salem. "Graph-driven analysis and visualisation of freebase schema as a directed weighted graph." *European Journal of Computer Science and Information Technology* 2, no. 4 (2014): 40–48.
- [18] Demidova, Elena, Iryna Oelze, and Wolfgang Nejdl. "Aligning freebase with the yago ontology." In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pp. 579–588. ACM, 2013.
- [19] Wang, Xue, Xuan Zhou, and Shan Wang. "Summarizing large-scale database schema using community detection." *Journal of Computer Science and Technology* 27, no. 3 (2012): 515–526.
- [20] <http://wiki.freebase.com/wiki/Schema>
- [21] <https://www.nuget.org/packages/Neo4jClient>
- [22] Shneiderman, Ben. "The eyes have it: A task by data type taxonomy for information visualizations." In *Visual Languages*, 1996. *Proceedings, IEEE Symposium on*, pp. 336–343. IEEE, 1996.
- [23] <http://freebaseviz.appspot.com>



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Towards a future mobile multihomed environment

A. Benaouda Chaht, C. Zouaoui & A. Bounoua
RCAM. Faculty of Technology, UDL, Sidi Bel Abbes, Algeria

ABSTRACT: Currently, network mobile nodes with multiple network interfaces, which support multiple network technologies, become the trend over the medium and long term. This requires a complete recast of the optimization mechanisms as well as the development of new designs for the support of mobility, in the context of the multi-homing. Therefore, the MobilityFirst is one of the bright projects proposed to address the challenges associated with the future mobile Internet. In this work, we present its features and discuss its impact on different application fields.

1 INTRODUCTION

Since a long time, the original and exceptional concept of IP generated a worldwide success of the Internet in all the practices of our society, whether in the economic, administrative, scientific, cultural and many other activities. The research community is constantly trying to improve the Internet and effectively meet the needs of contemporary uses. One of the primary needs is the availability of the Internet anywhere, even in case of failures. So the only method is to have a multi-access through multi-interfaces which means have multiple IP addresses, and this is the multi-homing [1]. In short, multi-homing aims to accomplish different goals [2]. First, resilience, as the diversity of multiple interfaces/paths can improve resilience since upon the failure of one interface/path, another one can be employed to provide connectivity. Second, ubiquity, since multiple network interfaces, in particular when used in a mobile and wireless network environment, enables ubiquitous access to the Internet over different media. Third, load sharing, as multiple interfaces/paths can be used simultaneously to improve throughput. Finally, flow distribution, as flows can be stripped in a dynamic way to meet user policies.

Supported by the latest technological progress, mobile terminals are equipped with multi-homing capability (e.g., the Samsung S5 smart phones can establish associations with both LTE and Wi-Fi networks using the Download Booster [4]) and enabled simultaneous access to different networks such as cellular networks (UMTS, HSDPA, LTE), wireless local area networks (802.11 family), and broadband wireless networks (LTE, WiMAX). Such as most of users prefer to be mobile, in that respect, we are faced with a need of new protocols that incorporates heterogeneous access networks for providing high quality mobile services.

In this paper, we explore the MobilityFirst [5] a clean-slate Internet architecture that addresses the challenges of wireless access and mobility at scale, while also providing significant improvements in performance and service quality, needed for emerging mobile Internet application scenarios.

The remainder of this paper is structured as follows. In session II, we discuss the approaches which propose to improve the performance expected by multi-homing in a mobile environment, where we think that the MobilityFirst project presents the best architecture to supporting the combination between mobility and multi-homing. Session III, we briefly overview the MobilityFirst functioning. An investigation of the main MobilityFirst works is presented in session IV, and Section V gives the concluding remarks.

2 DISCUSSION

Firstly, Multi-homing is leading to enhancements in well-known protocols, e.g. Mobile IPv6 (MIPv6) [6] which is, to a large degree, the archetypical mobility management protocol for IPv6 networks. For that purpose, an extension was standardized in RFC 5648 [7] Multiple Care of Address (MCoA) which extends MIPv6 to allow the registration of multiple addresses. Consequently, this extension allows a partial support multi-homing of end nodes and no site multi-homing e.g. if deployed in cars or trains.

Much prior research evaluating multi-homing support, are based on LIS concept such as architecture proposals in [8]-[9] which suggest separating the node identity ID, from the location, referred to as Locator. The ID is considered to be unique and independent of the Locator. Based on that, the Locator is used only for purpose of routing, while the unique ID is used only for persistent

node identification. This decoupling of functionalities, typically achieved an additional sub-layer at the end user node's network stacks that is used for mapping functionality between ID and Locator, helps to enhance the routing functionalities and provide mobility and multi-homing support. However, the mechanisms of the previous references mentioned above, are not fast enough to fulfill the requirements of real-time applications in mobile environments. Moreover, reference [12], was described to tackle the routing scalability issue and to enable site multi-homing. Nevertheless, this concept requires the usage of an additional mapping system and specialized border routers at a global scale. Moreover, this scheme may increase the latency of the core networks [13], which may negatively affect the Quality of Service (QoS) capabilities of the core networks.

Now, the most recent efforts of researchers have gone towards a transport layer protocol to multi-homing support, such as SCTP [14]–[16] and MPTCP [17]. The first one knew recent extensions to support the simultaneous use of several paths (multi-homing) [19]–[20]. Unfortunately, SCTP has not been widely deployed, except in niche applications such as signaling in telephony networks because, many firewalls and NAT (Network Address Translation) boxes are unable to process SCTP packets and thus simply discard them. On the other hand, MPTCP [21] is designed to resolve the problems of the previous protocol, and it is implemented in iOS 7 [22]. Authors in [23] perform extensive measurement based studies on MPTCP for dual-homed devices (with Wi-Fi and 3G/LTE).

While these end-to-end transport-layer proposals have started seeing some early deployments, it is not always possible to have a good end-to-end route to the destination in a wireless network with varying levels of connectivity.

Moreover, the requirements [26] for a robust Future Internet Architecture, push the National Science Foundation's [27] Future Internet Architecture (FIA) program to propose the Mobility-First architecture with the intention of addressing emerging mobility services and security needs from a clean-slate of the existing Internet. The key features of this architecture are:

- Name/Address Separation (NAS), where each device is provided with both a name called GUID (Globally Unique ID which is a public key name to secure user or device) and an address or addresses (in case of multi-homed devices). The address could be bound to the name any time during the transmission of a data packet. Thus, packets could be routed using either name or address. This late binding to an address enables the network to handle multi-

homing, where packets for a multi-homed device could be routed using only GUID;

- Routing Services; each node in the network uses in-network storage with hop-by-hop transport of large data units to deal with all forms of mobility and the associated challenges in a unified manner;
- Self-certifying public key network addresses to support strong authentication and security.
- A separate management plane which enables decentralized visibility of network resources and supports more general forms of service level agreements between network entities;
- An optional computing layer at the routers is introduced to enable service customization and security/privacy processing capabilities inside the network.

3 MOBILITYFIRST OVERVIEW

In this session, we show the main building blocks of the MobilityFirst design (see Fig. 1) and how they work together.

3.1 NAS (separation)

The main idea of the MobilityFirst is the separation of network-connected objects names' which can be managed, such as a smart-phone, a person, a group of devices/people, content or even context, from their locators (addresses) in order to ensure the ongoing mobile communications. In fact, GUIDs (Global Unique ID) assigned for network attached objects by a name certification service to the objects, they are basically public keys thereby enabling authentication and security services in the network; and used as the long-lasting network level identifiers for these objects. Thereafter, a dynamic mapping of GUIDs to one or more Network Addresses (NA) is made through a logically centralized, but physically distributed

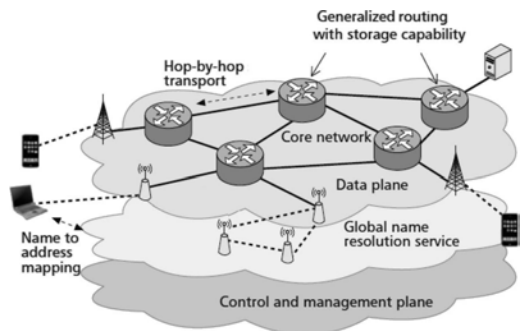


Figure 1. MobilityFirst architecture.

infrastructure called the Global Name Resolution Service (GNRS) [29]. This service re-lookups a destination IP address(es) during a change of a node network’s attachment but the GUID stay the same. Therefore, at an Internet scale, a router DHT-based Direct Mapping (DMap) [30] provides a fast global name resolution service for achieving a good balance between scalability, low update/query latency, consistency, availability and incremental deployment. This helps meet mobility and multi-homing requirements.

3.2 Routing services

MobilityFirst uses the Edge-Aware Inter-domain Routing (EIR) protocol at the inter-domain level for more information, refer [31]. And at a local scale (relatively close to end users), it uses the Generalized Storage Aware Routing (GSTAR) protocol [32] which is a proactive link state protocol, with added DTN [33] capacities in order to support disconnections and the critical conditions of mobile networks.

In particular, each GSTAR router maintain two types of topology information: (i) The intra-partition graph is formed by collecting Flooded Link State Advertisements (F-LSAs) which carry fine-grained, time-sensitive information about the links in the network; (ii) The second, termed the DTN graph, is maintained via epidemically Disseminated Link-State Advertisements (DLSAs) which carry connection probabilities between all nodes in the network [34]. Moreover, each router must have some storage capability and make forward vs. store decisions based on both short-term and long-term path quality metrics in order to provide improved performance in presence varying link quality and disconnection. This makes the basic idea in GSTAR which is an in-network storage routing to improve service quality and throughput in wireless access network. Finally, GSTAR combines a set of data packets into chunks (autonomous unit of message transmission) and provides hop-by-hop reliability for data packets transmission.

4 MOBILITYFIRST WORKS

The MobilityFirst protocol has been extensively validated using a combination of simulation, emulation and experimental trials on GENI platform [44].

In this session, we discuss about the different works addressing the MobilityFirst. Firstly, we remark that MF is a generic solution for cellular/Internet integration including new use cases such as heterogeneous network access, wireless peer-to-peer (P2P) networking, vehicular (V2V)

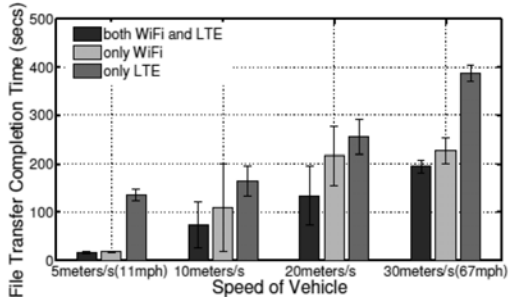


Figure 2. File transfer completion times for a multi-homed mobile client with a Wi-Fi and an LTE interface [43].

networking, Internet of Things (IoT) [35]–[37] sensor and machine-to-machine (M2M) applications, and so on. So in [38]–[40], the benefits of MF in this applications fields are already proven.

Moreover, the results in [41] indicate that by intelligently utilizing in-network storage, GSTAR outperforms traditional and storage-augmented link-state protocols in both wired and wireless network environments.

Furthermore, the recent studies of MobilityFirst’s team members, try to enhance their project. Consequently, the work presented in [43], proposes a specific data striping algorithm which allows simultaneous data transfer across multiple interfaces with per-flow based back-pressure link quality estimation and Figure 2 proves the performance expected by this proposition applied on vehicular nodes when they opportunistically use Wi-Fi hotspots while already being connected to an LTE network.

The most recently, D. Raychaudhuri & ray demonstrated mobility service with a dual-homed smartphone over the GENI network [45] with 12 MobilityFirst routers across the US and multiple wireless access networks with 4G/Wi-Fi service.

5 CONCLUDING REMARKS

In this paper, we present an overview of the MobilityFirst project, which has been extensively validated. This architecture is introduced as a clean-slate design with key features including name/address separation, storage-aware routing and hop-by-hop transport, robustness with respect to link quality variation and disconnection, content/context addressability. More importantly, the clean separation of names and addresses enables seamless mobility and enhances the use of multi-homing this makes from the MobilityFirst project the most potential approach to achieve the special

requirements of the emerging multihomed mobile devices.

REFERENCES

- [1] Hurson, A. 2012. *Connected Computing Environment*. San Diego: Elsevier.
- [2] Espi, J. Atkinson, R. Andonovic, I. & Dunlop, J. 2009. Proactive Route Optimization for Fast Mobile IPv6. In the *70th Vehicular Technology Conference (VTC 2009-Fall)*. Anchorage, AK: IEEE.
- [3] Launois, C.D & Bagnulo, M. 2006. The paths toward IPv6 multihoming Published in: *IEEE Communications Surveys & Tutorials* 8(2): 38–51.
- [4] Galaxy S5 Download Booster, <http://galaxys5.guide.com/samsung-galaxy-s5-features-explained/galaxy-s5-download-booster/>
- [5] MobilityFirst Future Internet Architecture Project, <http://mobilityfirst.winlab.rutgers.edu/>.
- [6] Johnson, D. Perkins, C. & Arkko, J. 2011. RFC: 6275: Mobility Support in IPv6. Available at <https://tools.ietf.org/html/rfc6275>. Access 12 Jan 2014.
- [7] Wakikawa, R. Devarapalli, V. Tsirtsis, G. Ernst, T. & Nagami, K. (2009). *Rfc 5648: Multiple care-of addresses registration*. Available at <http://tools.ietf.org/html/rfc5648>. Access 24 Jan 2014.
- [8] Nordmark, E. & Bagnulo, M. 2009. Rfc 5533: Shim6: Level 3 multihoming shim protocol for ipv6. Available at <http://tools.ietf.org/html/rfc5533>. Access 28 Mar 2014.
- [9] Moskowit, P. & Nikander, P. 2006. *Rfc 4423: Host identity protocol (hip) architecture*. Available at <http://www.ietf.org/rfc/rfc4423.txt>. Access 15 Mar 2012.
- [10] Moskowit, R. Nikander, P. Jokela, P. & Henderson, T. 2008. *Rfc 5201: Host identity protocol*. Available at <http://www.ietf.org/rfc/rfc5201.txt>. Access 15 Mar 2012.
- [11] Dhraief, A., & Montavont, N. 2008. Toward mobility and multihoming unification- the shim6 protocol: A case study. In *IEEE wireless communications and networking conference, WCNC*. 2840–2845.
- [12] Farinacci, D. Fuller, V. Meyer, D. & Lewis, D. 2013. Rfc 6830: The Locator/ID Separation Protocol (LISP), IETF Internet Standard. Access 15 Dec 2015.
- [13] Wang, Y., Bi, J., & Wu, J. 2010. Empirical analysis of core-edge separation by decomposing internet topology graph. In *IEEE Global Telecommunications Conference (GLOBECOM 2010)*. 1–5.
- [14] Wallace, T.D. Shami, A. 2012. A review of multihoming issues using the stream control transmission protocol. *IEEE Commun. Surv.Tutorials*. 14(2). 565–578.
- [15] Stewart, R. 2007. *Rfc 4960: Stream control transmission protocol*. Available at <http://tools.ietf.org/html/rfc4960>. Access 20 Sep 2013.
- [16] Dreibholz, T. Rathgeb, E. Rüngeler, I. Seggelmann, R. Tüxen, M. & Stewart, R. 2011. Stream control transmission protocol: Past, current, and future standardization activities. *IEEE Communications Magazine*, 49(4), 82–88.
- [17] Ford, A. Raiciu, C. Handley, M. Barre, S. Iyengar, J. 2011. *Rfc 6182: Architectural guidelines for multipath tcp development*. Available at <http://tools.ietf.org/html/rfc6182>. Access 04 Apr 2013.
- [18] Dreibholz, T. Becke, M. Rathgeb, E.P. & M. Tüxen 2010. On the use of concurrent multipath transfer over asymmetric paths in *Proc. of GLOBECOM*: IEEE.
- [19] Wallace, T.D. Shami, A. 2012. A review of multihoming issues using the stream control transmission protocol. *IEEE Commun. Surv.Tutorials*. 14(2): 565–578.
- [20] Iyengar, J. R. Amer, P. & Stewart, R. 2006. Concurrent multipath transfer using SCTP multihoming over independent end-to-end paths. *IEEE/ACM Transactions on Networking* 14(5): 951–964.
- [21] Wischik, D. Raiciu, C. Greenhalgh, A. & Handley, M. 2011. Design, implementation and evaluation of congestion control for multipath tcp. In *NSDI*.
- [22] Apple ios 7 surprises as first with new multipath tcp connections. <http://www.networkworld.com/news/2013/091913-ios7-multipath-273995.html>.
- [23] Chen, Y. Lim, Y. Gibbens, R.J. Nahum, E.M. Khalili, R. & Towsley, D. 2013. A measurement-based study of Multipath TCP performance over wireless networks. in *Internet Measurement Conference*: ACM.
- [24] Han, H., Shakkottai, S., Hollo, C. V., Srikant, R., & Towsley, D. 2006. Multi-path tcp: A joint congestion control and routing scheme to exploit path diversity in the internet. *IEEE/ACM Transactions on Networking*, 14, 1260–1271.
- [25] Zhang, M. Lai, J. Krishnamurthy, A. Peterson, L.L. & Wang, R.Y. 2004. A transport layer approach for improving end-to-end performance and robustness using redundant paths. *USENIX Annual Technical Conference, General Track*.
- [26] Pan, J., Paul, S., & Jain, R. 2011. A survey of the research on future internet architectures. *IEEE Communications Magazine*, 49(7), 26–36.
- [27] NSF Future Internet Architecture Project, <http://www.nets-fia.net/>.
- [28] Seskar, I. Nagaraja, K. Nelson, S. & Raychaudhuri, D. 2011. Mobilityfirst future internet architecture project. In *Proceedings of the 7th Asian Internet Engineering Conference, ser. AINTEC '11*. New York, NY, USA: ACM.
- [29] Venkataramani, A. Sharma, A. Tie, X. Uppal, H. Westbrook, D. Kurose, J. & Raychaudhuri, D. 2013. Design Requirements for a Global Name Service for a Mobility-Centric, Trustworthy. *Internetwork. Fifth International Conference on (COMSNETS)*, IEEE: 1–3.
- [30] Vu, T. et al. 2012. DMap: A Shared Hosting Scheme for Dynamic Identifier to Locator Mappings in the Global Internet. in *Proceedings of ICDCS*.
- [31] Vu, T. Baid, A. Nguyen, H. & Raychaudhuri, D. 2012. EIR: Edge-aware Interdomain Routing Protocol for the Future Mobile Internet, *WINLAB Technical Report, WINLAB-TR-414*. (PDF)
- [32] Nelson, S.C. Bhanage, G. & Raychaudhuri, D. 2011. GSTAR: Generalized Storage-Aware Routing for MobilityFirst in the Future Mobile Internet. in *Proc. of MobiArch*: ACM.

- [33] Ott, J. Kutscher, D. & Dwertmann, C. 2006. Integrating DTN and MANET routing. In *Proc. of ACM CHANTS*.
- [34] Whitbeck, J. & Conan, V. 2010. HYMAD: Hybrid DTN-MANET routing for dense and highly dynamic wireless networks. *Computer Communications Journal*. 33(13): 1483–1492.
- [35] Atzori, L. Iera, A. & Morabito, G. 2010. The Internet of Things: A survey. *The International Journal of Computer and Telecommunications Networks* 54(15): 2787–2805.
- [36] Bandyopadhyay, S. Sengupta, M. Maiti, S & Dutta, S. 2011. A Survey of Middleware for Internet of Things. *Recent Trends in Wireless and Mobile Networks*: Springer Berlin Heidelberg.
- [37] Li, J. Shvartzshnaider, Y. Francisco, J.A & Martin, R.P. 2012. Enabling Internet-of-Things services in the MobilityFirst Future Internet Architecture. *International Symposium on World of Wireless, Mobile and Multimedia Networks*, IEEE.
- [38] Mukherjee, S. Baid, A. & Raychaudhuri, D. 2015. Integrating advanced mobility services into the future Internet Architecture. *7th International Conference Communication Systems and Networks (COMSNETS)*: IEEE.
- [39] Baid, A. & Raychaudhuri, D. 2012. Wireless access considerations for the MobilityFirst future Internet architecture. *The 35th IEEE Sarnoff Symposium (SARNOFF)*.
- [40] Li, L. Zhang, Y. Nagaraja, K. & Raychaudhuri, D. 2012. Supporting efficient machine-to-machine communications in the future mobile internet. *Wireless Communications and Networking Conference Workshops (WCNCW)*: IEEE.
- [41] Somani, N. Chanda, A. Nelson, S.C. & Raychaudhuri, D. 2012. Storage-Aware Routing for Robust and Efficient Services. *International Conference on Communications*: IEEE.
- [42] Raychaudhuri, D. Seskar, I. Ott, M. Ganu, S. Ramachandran, K. Kremos, H. Siracusa, R. Liu, H. Singh, M. 2005. Overview of the ORBIT Radio Grid Testbed for Evaluation of Next-Generation Wireless Network Protocols *Wireless Communications and Networking Conference*: IEEE.
- [43] Mukherjee, S. Baid, A. Seskar, I & Raychaudhuri, D. 2014. Network-assisted multihoming for emerging heterogeneous wireless access scenarios. *25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*, IEEE.
- [44] Raychaudhuri, D. 2013. Using GENI to Prototype the MobilityFirst Future Internet Architecture. <http://www.geni.net/?p=2644>.
- [45] Raychaudhuri, D & ray. 2015. MobilityFirst: A Clean Slate Network Architecture for Next-Generation (“5G”) Mobility Services. *Next-Generation Mobile Network Architecture*.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Information security challenge: The responsibility of management, Information System case study for the management of research

Y. El Hissi

Computer, Networks, Mobility and Modeling Laboratory, Department of Mathematics and Computer FST, Hassan 1st University, Settat, Morocco

A. Haqiq

Computer, Networks, Mobility and Modeling Laboratory, Department of Mathematics and Computer FST, Hassan 1st University, Settat, Morocco
e-NGN Research Group, Africa and Middle East

ABSTRACT: The context of current governance necessitates the collection, processing and sharing of information allowing organizations to develop their methods of management and decision making.

Managing the process of scientific research in the Moroccan university uses the information technology to promote its activities and enhance its performance.

We must be certain of tools and elements that provide us with that information, means that allow its conservation and protection and ensure the channels involved in its dissemination and communication.

Information System as a technology carrier compound of this information, it must be well protected and secured for any technical threat, physical or logical.

And so, this article deals with the research and design of tools and reliable and effective means for securing the Information System used in the scientific research activity; if necessary, the management system of information security based on ISO 27001. This in order to meet the management requirements, communication and security of scientific research in Moroccan universities.

Keywords: Information, Information System, decision making, Moroccan university, scientific research, security, system management, governance, ISO 2700 standard

1 INTRODUCTION

Information is the base of knowledge. Its value is therefore materialized on the creation of the information itself and it develops through the exchange and sharing.

This information is most effective when it helps the user to achieve its objectives, it is especially important for the activity of scientific research where it is considered raw material available or because the information asymmetry in decision making games requires seeking additional information to achieve an acceptable level of individual knowledge to decide.

The collection of information demand to gather all the internal and external competences of an organization to better identify the needs and the use of research methods, analysis and dissemination of information.

Wrong information, falsified or weakened lost its value and influences the quality of deliverable.

On scientific research information is an important factor as the human, material, financial and social for the success of the activity; and if we

multiply the tools and means to control risks associated with these factors, we must lead the same policy to secure information.

Since the Information System (IS) represents all the elements involved in the management, storage, processing and sharing of information, security becomes a necessary way to lend credibility to its value, its effectiveness and its purpose.

In the field of higher education in Morocco, the use of new technologies has become an unavoidable reality, since its adoption in national projects; this is an important factor in the promotion of its activities, the value of its results and their communication with all stakeholders.

Being convinced that this type of information still being implemented in Morocco is sensitive; its security is needed to protect its shape and value at every stage of scientific production until it becomes public information. This is because information security is the operation indented for protection of information against a large cluster of threats in order to ensure business continuity (Mataracioglu, Yildirim and Bilgem 2014).

It is generally accepted that Information Security Governance is an integral part of Corporate Governance (Solmsa, Basie & Solmsb 2006).

This governance of information security makes sense since it consists to establish a structure to make the right security decisions at the right time at the right hierarchical level.

So, in this article we will look for effective and efficient means for securing the IS used in the activity of the research. This will be mainly based on the identification and assessment of risks to information on scientific research, analysis of operation of research structures, the study of existing tools and proposal opportunities to master the safety of this information.

2 INFORMATION SYSTEM SECURITY “ISS”

Risk is the “effect of uncertainty on objectives” and an effect is a positive or negative deviation from what is expected (Saporita, Sciuttob and Sciuttoc 2014).

A risk is a possible danger, more or less predictable, inherent to a situation or activity. Risk means also a dangerous uncertainty, shorthand summary of the longer definition as the expectation of negative consequences of current or potential, future phenomena or events (Zartman 2009).

A risk is defined by a possibility of a future event, uncertain, which does not only depend on the willingness of the parties, and that can cause a loss or any other damage.

Risks are classified by their nature into one of three categories:

- External risks, related to the environment of the organization, its business, its search field, the host university policy regulations, etc.
- Internal risks related to the body, the research unit, its management, its processes, information systems, etc.
- Steering risk, related to information needed to make good decisions: Dashboards

In our study we focus on internal risks and trying to find ways to control them.

In this category of risks, risks related to information have its own characteristics. Their factors are manifold:

- Technology: malfunction of a component in a technical or application infrastructure, can disrupt the supply of a service, causing loss of confidentiality of information or harm the integrity of the information assets of the research unit,
- Human: computer crime, hacking, spying, human error in the selection or the use of an IT solution,
- Natural, primarily climate risks: heat/cold, floods....

The Information Technology (IT) is a particular axis of risk management. Management of IT risk should ensure the availability, integrity, confidentiality of corporate data, as well as evidence and control.

If IT-related risks can be treated with the same approach as other risks (products, social and financial risks), but they must be analyzed and managed on the basis of close cooperation with the various activities of the Research Unit.

The major risks related to IS are:

- risk of failure, a major source of operational risk;
- risk of penetration of the organization by (eg espionage);
- risk of external attacks.

IT governance should be able to assist in risk control, that is to say, seek to preserve the value gained by the research unit against all deviations that could cause its depreciation or destruction.

Information security is primarily a management and business issue (Marnewicka and Labuschagne 2011) and securing IS, is an approach based on two issues: organization and technology, hence the importance of knowing the specific risks, perimeter security, identify the costs and benefits in kind.

Risks related to the IS can be complex to identify and manage; it requires action from the project development phase. Information security refers to defending information from unauthorized access, disclosure, use, modification, disruption, inspection, and perusal. In other words, confidentiality, integrity and reliability of information are important in information security (Soomro, Shah, and Ahmed 2016).

To do this, organizations need to understand the methods, standards and solutions that will help optimize the information system performance while controlling risks.

There is enough method for information security and risk analysis as: MEHARI (Lizcanao, Azcorrab, Solé-Paretac and Pascualc 1999), EBIOS (Abdallah, Yakymets and Lanusse 2015) MELISA (Eloff, Labuschagne, Badenhorst 1993)... but in this article we focus on approaches based on issues management and provides consistent and organizational approaches, because it is necessary to go beyond technical considerations and adopt principles and organizational values that are going in the same direction of governance of IS.

We should know that Protection of information resources and technology; ensuring safe and secure processing of information; and ensuring reliable and safe flow of information is essential for both production and protection in intensive organizations (Albrechtsen 2015).

3 SCIENTIFIC RESEARCH AT MOROCCAN UNIVERSITY: NATURE OF INFORMATION, RISKS, OPPORTUNITIES AND SECURITY VULNERABILITY

Over the years, the development of higher education and the promotion of research were the target of national government projects, which aim to have rewarding research activities, innovative and meet the economic needs of the country and therefore influence life of the citizen.

Among these projects we mention:

- The Emergency Plan (2009–2012) (Ministry of National Education, Higher Education, Staff Training and Scientific Research Emergency Plan 2009–2012, November 2011) that gave birth to several actions such as Project 14 to the promotion of scientific research which introduced significant measures to improve the governance and monitoring of scientific research, enhancing the attractiveness of the research profession, increase, diversification and sustainability of funding sources for scientific research and exploitation of research.

In addition, every effort will be made to promote international cooperation in scientific research, as well as vehicle financing that emulation.

The strategy of the Corporate Plan for 2013–2016 (Departmental action plan for 2013–2016, March 2012): came to strengthen and continue the process of the Emergency Plan with 39 projects spread over 6 axes and designed all good management of the sector, the promotion of scientific research, improving the supply of education and social benefits to students, revision of laws governing the sector and also the development of a strong international cooperation strategy.

Projects that are interested in the management and promotion of scientific research are:

- PROJECT 4 Axis-II: the location of functional blocks of the information system, among its objective is the facilitation of obtaining information on scientific research and this through the establishment of system information for scientific research
- PROJECT 2 Axis-III: structuring of scientific research and that is the aim of integrating into the socio-economic environment and meet their needs and also to encourage excellence in scientific research.

To responds to the Measures Introduced by national projects for the advancement of research to enhance thesis must efforts use a technological device that organizes all the research activities that structure units and communicating the work

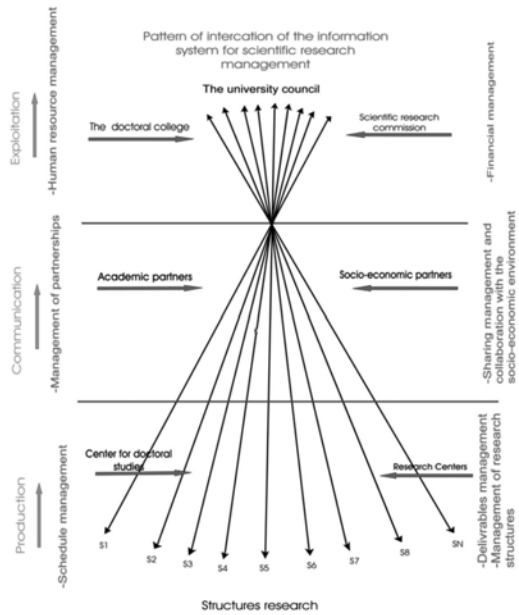


Figure 1. Synthetic scheme of interaction in the IS for management of scientific research.

to goshawks Enhance the sharing and opens a window and partnership with its socioeconomic environment.

The figure below shows clearly the objective expected by the implementation of the said IS:

- It is therefore necessary to establish a system for information security related to scientific research of all types of risks that can affect its value and power.

4 INFORMATION SECURITY MANAGEMENT SYSTEM

Managing security for information assets is a critically important and challenging task (Derek and Nazaretha 2015).

Information security management needs a paradigm shift in order to successfully protect information assets (Jan, Eloff and Eloff 2003).

In the ISMS, the information is not limited to computer systems. The information is in the broad sense. It must be studied in all its forms regardless of its support, human, paper, software, etc...

The term security must be understood as all the resources deployed to protect against malicious acts.

The big question being asked on information security, is "how to integrate standalone security solutions under a common framework and integrate with the strategic objectives of the organization".

The information security community has responded to this issue by developing standards that can be used to develop a safety management framework to aggregate information.

Organizations have begun to respond to these challenges by implementing technical measures to protect the channels of information and storage devices.

Although they were sufficient initially, as long as companies have evolved subsequently, new threats have emerged, where awareness has shifted the security visualization center of information of a solution technical to organizational management culture motivated (Sotaro, Baba and Walsh 2014).

IS security cannot be in any way the responsibility of technicians and/or IS leaders, it is a responsibility of leaders and managers in the first place, leaders must therefore be regarded as the starting point of a formal instead of ISS that can be considered satisfactory, making their important actions. The involvement of the leader is, also considered crucial in the establishment, maintenance and success of actions related to IS security. The importance of the involvement and action leader for ISS was demonstrated not only in literature but also in many guides to good practice, or in standards dedicated to the ISS, such as ISO 270xx1 (Shing., Yenb, Chenc, Chend., Wen and Chien-C.C 2015).

When we talk about management system, fantasies and anxieties come back, and this is also true in the field of information security.

Management systems are common reference of formalization process to "... guarantee certain characteristics of products and services such as quality, environmental friendliness, safety, reliability, efficiency and interchangeability".

With the development of information security standards, there is a "safety certification".

The British Standards Institute was the first organization in 1995 to publish a standard in the field of computer security. The BS 7799 standard defined the best practices for computer security (Germainn 2005).

The ISO (International Organization for Standardization) has followed five years later, and has published numerous standards in the same field, such as ISO 17799, BS end of 7799 or ISO 13335 (guidelines for management Security) (Shing., Yenb, Chenc, Chend., Wen and Chien-C.C 2015).

These standards are designed to ensure security of information; its support is paper or electronic nature and the cause of potential incidents either accidental or deliberate. It is then the "family" 2700x which develops the concept of "Information Security Management System" (ISMS). An ISMS provides a framework for continuous improvement of information security, based primarily on a risk management approach. At the moment, eight



Figure 2. Hierarchical structure of the working groups and committees of the ISO/IEC.

standards are under development in the 2700x series, some of which were published (including ISO 27001 defines the requirements for the certification of ISMS) [15]. Eventually, the integrated set of standards from the series 2700x expected to form a governance model for information security (Shing., Yenb, Chenc, Chend., Wen and Chien-C.C 2015).

It proposes a model that meets the challenges of an optimized and sustainable governance of information security. It streamlines and credibility to the security approach based on four principles: Leading from the risks, the process approach, Management involvement and continuous improvement is simply the principle of PDCA (Micića, Micićb and Blagojević M., 2013).

The ISS ensures that the safety devices are refined to keep pace with changes in security threats, vulnerabilities and business impacts, which considered an important aspect in a dynamic field and one of the main benefits risk flexibility oriented approach ISO27k (Shing., Yenb, Chenc, Chend., Wen and Chien-C.C 2015).

ISO 27001 adopts an approach of compliance/non-compliance based on a coherent approach of PDCA can be applied to any organization, whatever its conditions and regardless of their environment and risks.

And its implementation in any organization as basic Management System Information System that helps streamline and using a credible security on one hand and the growth continues to improve its approach.

5 CASE STUDY: SECURITY MANAGEMENT INFORMATION RELATED RESEARCH IN MOROCCAN UNIVERSITY BY THE ESTABLISHMENT OF THE STANDARDS ISO 27001: MANAGEMENT SYSTEM OF INFORMATION SECURITY "MSIS"

As the Moroccan university is integrated into projects being implemented information systems for various brick business such as human resource

management, accounting, assets ...; the research management also plays an important part in the university's development strategies (Ministry of National Education, Higher Education, Staff Training and Scientific Research, "Morocco's strategy for the development of Research 2025", Science Branch—November 2009).

An information system as developed, allows us to manage a wealth of information, a set of data and interactions; in short it is a real asset for the university to ensure a comprehensive research management.

One such tool is a primary device that will provide a significant improvement to the development, communication and enhancement of research at the university level. Therefore it is necessary to ensure the effectiveness and performance of the information system; this can be achieved by developing a management and security system that allows both organizing and securing the SI established.

This management and security system must ensure AIP three objectives:

- The Availability: access to information at the right time
- Integrity: Banning any stored information
- The Privacy: prohibition on access to sensitive information to unauthorized persons.

These three security principles can be heard; the ISMS include other concepts such as authentication, traceability, non-repudiation, accountability which are security mechanisms are deployed based on security needs of the body.

ISO 27001 management system deployed in the Information Security "ISMS" is the approach that meets all these needs and provided a set of requirements that will facilitate the management of information including those related to data personal, the financial data, information on future projects and even to documents subject to intellectual property.

ISO 27001 provides a model that meets the challenges of an optimized and sustainable governance of information security, it is based on four key principles:

The choice of ISO 27001 as a repository for the Security of the Information System in place for

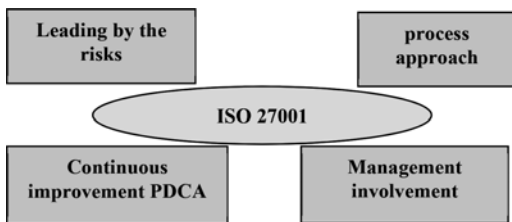


Figure 3. Information security governance model.

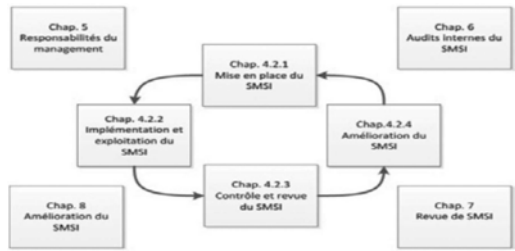


Figure 4. Structure of ISO/IEC 27001.

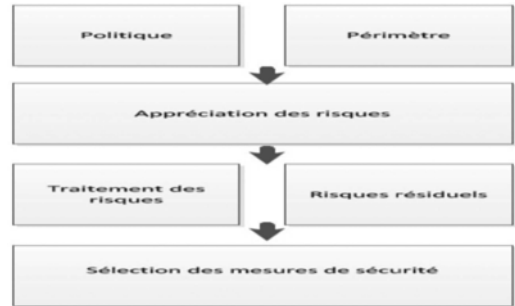


Figure 5. Steps phase plan PDCA.

the management and communication of research in the Moroccan university, lies in the made that the standard can be integrated easily and completely with other standards in the family and the Moroccan university began to adopt and implement as its strategies and orientations.

All of these standards have the same structure which facilitates their integration thereafter to an integrated management system.

To be in accordance with ISO/IEC 27001 standard, the MSIS must meet all requirements between chapters 4 and 8 shown in the following figure (Shing., Yenb, Chenc, Chend., Wen and Chien-C.C 2015).

Chapter 4 is the center of the standard, it includes the four phases of the PDCA (PLAN, DO, CHECK, ACT) Deming, we will follow step for the implementation of the MSIS.

5.1 Phase plan: Planning the process of securing IS

This phase is to set the MSIS goals by following four steps:

5.1.1 Policy and the scope of the ISMS

The security policy for the management of the research is introduced to specify the level of security that will be applied within the scope of the ISMS. The standard does not set requirements on the perimeter, it may be restricted or cover all

activities related to scientific research at the university. The aim is to include the activities for which the stakeholders in research requires a certain level of confidence.

5.1.2 Risk assessment

The assessment of risks related to scientific research can be treated by many methods as EBIOS or Mehari.

ISO/IEC 27001 it only sets specifications specifying each of the key steps of risk assessment. The university is free to choose, develop his own method following the objectives set by the ISO/IEC 27001 or to apply an already proven, in order to identify the risks associated with the management and communication of research.

First, we must make a list of all assets that have an information material importance in scientific, in the ISMS. There are generally six categories:

- Material assets for all equipment, network and system.
- Physical assets, for offices, scientific production places from start to publication of results.
- Software assets, databases, files, and operating systems if any
- Human assets for all stakeholders involved in the management and communication of research as teachers, researchers, doctoral students and leaders in the field.
- Active Documents, for paper documents, user manuals.
- Intangible assets, to the knowledge of the university's scientific output.

Thereafter must be assigned for each information asset "owner". This is the person who best knows the value and consequences of compromise in terms of availability, integrity and confidentiality of the assets.

Then there is the step of identifying vulnerabilities of assets listed above. Vulnerability is the intrinsic property of the property that exposes threats. For example, a laptop computer is vulnerable to theft but the vulnerability is not theft but its portability. In this case the identification of the vulnerability is portability. It takes after identifying threats to the so-called asset information and assess the impact of loss of confidentiality, availability or integrity of the assets.

5.1.3 Risk treatment

Standard treatment options identified four risk:

- «Accept» the risk, returns to not deploy any security measure other than those already in place.
- «To avoid» the risk, is to remove such activity or material offering a risk.

- «To transfer» a risk by buying insurance or outsourcing.
- «Reduce» the risk, is to take technical and organizational measures to reduce to an acceptable level the risk. It is the most common treatment.

5.1.4 Selecting security measures

This step is to select the security measures. ISO/IEC 27001 provides 133 measures safety spread over eleven chapters. At this stage, the work is to develop a picture called SoA (Statement of Applicability) that lists the 133 measures should be reported applicable or not applicable, to reduce the risk of MSIS (Shing., Yenb, Chenc, Chend., Wen and Chien-C.C 2015).

5.2 Phase Do: Up implementation of security measures

This phase is the heart of the norm, since it allows to organize a plan for the measures to implement in practice. For this he must go through the selection of measures to secure the SI Search, identification of performance indicator to test the effectiveness of security measures and indicators to monitor the compliance of the ISMS and the training of personnel can begin with a reminder of their commitment to safety business and continue with a list of tips such as compliance with certain safety rules for passwords and working environment, and finally maintenance of the ISMS the procedures of ensuring the proper functioning of each of the ISMS process and ensure that their documentation is up to date.

5.3 Phase Check: Implementation of control means

An ISMS in place should be systematically monitored to measure its effectiveness and compliance, and therefore the security of the SI-search process will be subject to verification tools, such as:

- Internal audits: review previously planned involving auditors;
- Internal controls: controls within the university to verify the application of the procedures put in place permanent;
- Reviews: Always check that the ISS is in line with the particular environment of the university and adjust it each time.

5.4 Phase Act: Implementation of action

After assets established monitoring tools and detect malfunctions of the ISMS to the SI-Recherche, it is important to put in place: corrective actions, preventive and improvement actions to propose

action that allow improved the performance of the security of the IS-research.

Results of the different actions should be recorded and communicated to stakeholders of scientific research. These actions contribute to more effective and efficient MSIS.

6 CONCLUSION

In a globalized world, information has become a commodity of knowledge and expertise and we are each user or information provider.

The effectiveness of information is that it helps to make the decision, and therefore its value and credibility directly impacts this decision.

The IS system information since the whole element to the collection and sharing of information, it is an aid to decision-the decision reducing the level of uncertainty.

The quality of the decision is based primarily on the quality of SI and consequently the performance of the organization.

This is why the security of the information system is necessary for the collection and dissemination of sound and efficient information.

So as part of the governance of information security, the implementation of a management system for managing information security can be most effective through the intermediary of ISO 27001 which is the most common and the most referential overall.

Thus in the context of operation of the Moroccan universities and research units, the implementation of a management system for managing information security can be most effective through the intermediary of the standard ISO 27001.

REFERENCES

- Abdallah, R., Yakymets, N., Lanusse, A., 2015, Towards a model-driven based security framework, A Modelsward 2015, 3rd International Conference on Model-Driven Engineering and Software Development, Proceedings, 2015, 639–645.
- Albrechtsen E., 2015, Major accident prevention and management of information systems security in technology-based work processes, *Journal of Loss Prevention in the Process Industries*, Volume 36, July 2015, 84–91.
- Departmental action plan for 2013–2016, March 2012, p. 36.
- Derek L., Nazaretha, J.C., 2015, A system dynamics model for information security management, *Information & Management*, Volume 52, Issue 1, January 2015, 123–134.
- Eloff J.H.P., Labuschagne L., Badenhorst K.P., 1993, A comparative framework for risk analysis methods, *Computers & Security*, Volume 12, Issue 6, October 1993, 597–603.
- Germain R.S., Information Security Management Best Practice Based on ISO/IEC 17799, *The Information Management Journal*, July/August 2005.
- Jan H. Eloff P., Eloff M., 2003, Information security management: a new paradigm, Proceedings of the 2003 annual research conference of the South African institute of computer scientists and information technologists on Enablement through technology, 130–136.
- Lizcano P.J., Azcorra A., Solé-Paretac J., Pascual J.D., 1999, MEHARI: a system for analysing the use of the internet services, *Manuel Alvarez-Campanad, Computer Networks*, Volume 31, Issue 21, 10 November 1999, 2293–2307.
- Marnewicka C., Labuschagne L., 2011, An investigation into the governance of information technology projects in South Africa, 1, *International Journal of Project Management*, Volume 29, Issue 6, August 2011, 661–670.
- Mataracioglu T., Yildirim S. O. & Bilgem T. 2014, Obstructions of turkish public organizations getting ISO/IEC 27001 certified, *international journal of managing value and supply chains*, vol.5, No. 2, June 2014.
- Micića Z., Micić M., Blagojević M., 2013, ICT innovations at the platform of standardisation for knowledge quality in PDCA, *Volume 36, Issue 1, November 2013*, 231–243.
- Ministry of National Education, Higher Education, Staff Training and Scientific Research Emergency Plan 2009–2012, November 2011.
- Ministry of National Education, Higher Education, Staff Training and Scientific Research, “Morocco’s strategy for the development of Research 2025”, Science Branch—November 2009, 9–12.
- Saporita E., Sciuttob M., Sciuttoc G., 2014, Quantitative Approach to Risk Management in Critical Infrastructures Transportation Research Procedia Volume 3, 2014, 17th Meeting of the EURO Working Group on Transportation, EWGT2014, 2–4 July 2014, Sevilla, Spain, 740–749.
- Shing H.L., Yenb D.C., Chenc S.C, Chend P.S., Wen W.L., Chien-C.C., 2015, Effects of virtualization on information security, *Computer Standards & Interfaces* Volume 42, November 2015, 1–8.
- Solmsa R.V., Basie S.H. & Solmsb V., 2006, Information Security Governance: A model based on the Direct-Control Cycle, *Computers and security*, Volume 25, Issue 6, September 2006, 408–412.
- Soomro Z.A., Shah M.H., Javed Ahmed J., 2016, Information security management needs more holistic approach, *International Journal of Information Management*, Volume 36, Issue 2, April 2016, 215–225.
- Sotaro S., Baba Y. & Walsh J.P., 2014, “Organizational Design of University Laboratories: Task Allocation and Lab Performance in Japanese Bioscience Laboratories”, December 2014.
- Zartman W., Risk and Prevention in Identity Negotiations, 2009 *De Boeck Supérieur, Negotiations* 2009, n° 11, 77–92.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Effects and impact of playing computer games

W. Chmielarz & O. Szumski

Faculty of Management, University of Warsaw, Warsaw, Poland

ABSTRACT: The main aim of this article is to present the effects and impact of playing e-games. In order to realize the following aim, the authors conducted surveys limited to a selected group of individual users. In the paper the authors presented the findings based on the opinions of e-gamers concerning the social effects of taking part in games (e.g. the prestige resulting from playing computer games), psychological and physical effects of playing games, the approach towards participating in games, taking into consideration such elements as entertainment, sport and hobbies connected with playing games, or otherwise related to gaming. The authors have held a discussion of the obtained results and they have drawn conclusions based on the present stage of the research.

1 INTRODUCTION

The main aim of this work is to analyze the use of computer games as one of the alternative forms of entertainment in the selected group of users under the circumstances of a dynamic development of devices and mobile applications running on them. The aim of this article is to analyze the situation where computer games are used by people who treat them not only as a form of entertainment but also as a kind of sport. The popularity and specific universal nature of the access to computer games facilitates a fast development of information technologies. A broadly defined concept of mobility also impacts the use of computer games, moving the focus from using PCs to the use of smartphones and tablets.

According to the statistics of Newzoo service (GRY-OnLine 2014), in Poland in 2013 the number of gamers amounted to 13.4 million, out of which 98% used their PCs to play computer games (together with other platforms). We take the second position in Europe among the examined countries. The market of computer games in Poland is growing every year—in the end of 2014 it was worth about 280 million dollars and it will be growing by 3.8% a year, thus increasing the value of the entire market to 437 million dollars at the end of 2016 (Akcyjariat Obywatelski 2013). Hence, undoubtedly the subject matter is worthy of attention.

Unfortunately, the phenomenon itself is difficult to define and examine taking into account the formalized scientific analyses. Firstly, there is no clear definition of computer games (Chmielarz, W. 2015a, GRY-OnLine 2004, IT-Pomoc.pl 2016, KIPA 2016, PTBG 2010, Wiedza i Edukacja 2009, Wikipedia 2016, Zając J. 2014a). In its

narrow sense, this concept is treated literally as games in the form of software running only on traditional hardware such as (desktop, microcomputers, laptops or palmtops). In its broad, historical approach, the group encompasses also games running on devices such as a console, TV, gaming machines, smartphones and tablets (which are in fact communication and application computers). As the games running on all kinds of devices were being developed in parallel, and, in fact, there are PC equivalents of all kinds of games, we sometimes use this term in its broad meaning. Thus, for the needs of this study, the authors assumed that computer games are a generic term (hypernym) encapsulating the whole class of all kinds of games presented as a homogenous phenomenon. Secondly, there is no one generally accepted definition of a person playing computer games (e-gamer). Thus, in the narrow sense of the word, an e-gamer is a person who plays computer games every day or a few times a week, individually or taking part in a multi-player game. Sometimes, the scope of this term is limited to include only those players who treat MMO class games as a sport, and they try to play them professionally. However, we observe a more and more common tendency to expand the term to include also any individuals who play any kind of game from time to time, perceiving it as just one more alternative kind of entertainment. This article treats the concept of e-gamers in such a way. Thirdly, there is no (specific or clear) classification of computer games: there are a number of typologies based on various criteria, most frequently taking into account the type of activity required from the e-gamer playing games (e.g. logic, strategic, arcade, RPG, etc. games), with a number of varying kinds and versions.

The phenomenon of computer games has been examined in numerous studies (Mijal M. & Szumski O. 2013, Żywiczyńska E. 2014a), including large-scale studies (Żywiczyńska E. 2014b); nevertheless, they were carried out before the recent period of extreme popularity and growth in the number of applications running on smartphones and tablets. The authors hoped to establish certain implications of the new phenomena with regard to the direction of computer games development. Therefore, the authors have undertaken the studies whose main aim is to analyze the use of such applications among users. The findings presented in this article constitute a brief report on the first stage of the research conducted among the gamers in Poland in 2015.

2 THE ASSUMPTIONS OF RESEARCH METHODOLOGY

Due to limited and fragmentary research concerning the area of internet computer games and e-gamers, both from the point of view of an individual client and a group of customers, in Polish and foreign literature, the studies have been based on the authors' own approach (Chmielarz, W. 2015b) consisting of the following steps:

- analysis of a selected group of players on the basis of a quantitative and qualitative survey, divided into the following parts:
 - characteristics of a computer player and identifying his or her preferences in computer games,
 - identification of potential effects and consequences of playing computer games for e-gamers.
- placing an internet version of a survey on the servers of the Faculty of Management of the University of Warsaw, conducting functionality test and its verification,
- carrying out the survey among the users, analysis and discussion of the findings,
- drawing conclusions from the obtained results concerning the current situation and possible directions of the future development of internet computer games on the basis of the users' opinions.

The article presents the results of the analysis of the first part of the completed survey. It allowed for identifying a particular group of people who play various kinds of games, using different kind of hardware and software, with a varying level of skills and expectations concerning the organizational and technical aspects of playing games. Only after the selecting the group of best, “professional” players, we may proceed to specify the implica-

tions and psychophysical effects of their involvement in individual and multi-player games. The latter aspect was examined in the second, sequentially conducted, stage of the survey, whose results and conclusions will be presented in subsequent publications.

The questionnaire surveys were conducted near the end of December 2015. The selection of the study sample was not accidental: it belonged to the category of convenience sampling, the respondents were mainly students of selected universities in Warsaw (University of Warsaw and Vistula University (Akademia Finansów i Biznesu Vistula)), of full-time and part-time BA, BSc and MA studies. The survey was also completed by two members of university staff who declared playing computer games. The surveys were circulated electronically, and the response rate did not exceed 70%. Students are particularly open to all kinds of innovation, especially if it concerns their private life or entertainment.

A specific limitation concerning this particular sample was an anticipated high percentage of smartphone, tablet, laptop and mobile phone users, devices of lower quality but with a longer durability. The survey was completed by 274 people, out of which 254 participants submitted correctly completed questionnaires (which constitutes 92.70% of the sample). Among the respondents there were 59.45% of women and 40.16% of men; 0.39% respondents did not answer this question. An average age of the respondent was 20.62 years, and the medium value was 19 years. The age is typical of students of the first years of BA and BSc students and the first years of the studies of the second cycle—the group asked to complete the questionnaires. The oldest person taking part in the survey (member of the university staff) was 37. Among the survey participants there were 63.39% of students, 35.83% working students and 0.79% employees. 70.87% indicated secondary level education and 20.08% post-secondary education—the survey was primarily conducted among the students of BA studies. 8.66% declared holding a BA degree or a certificate of completion of studies, only one person indicated having a PhD degree.

Over 45% of survey participants indicated that they are inhabitants of cities with over 500,000 residents, over 14% came from cities with 100,000–500,000 of inhabitants, over 21% from towns with 10,000–100,000 residents, almost 5% from towns up to 10,000 residents, and 12.6% declared that they come from rural areas. The simplicity of the survey did not cause many distortions during its completion; few respondents (17) completed also additional sections of the survey.

3 THE ANALYSIS OF THE OBTAINED RESULTS AND RELEVANT DISCUSSION

The respondents have provided responses to forty-one substantive questions. The answers to the last twenty questions concerned the issues which are directly related to the objective of the present article. The first group of questions concerned the characteristics of e-gamers and the scope of the use of computer games. On this basis the authors formulated more difficult questions concerning the effects and consequences of participating in computer games.

Nearly 20% of respondents provided positive answers to the question concerning the greater prestige of a particular gamer among friends or acquaintances outside the game which he or she currently plays (Fig. 1). It is not the score which would give evidence to the wide influence of this form of entertainment, or its particular importance for the circle of gamers' friends. In general, the present results are similar to the responses obtained in the case of the second question as regards the formation of a circle of friends made up of other gamers, who play the same games at a particular moment (Fig. 2). A slight difference which amounts to about 6 percentage points probably results from the fact that, as reflected in the responses in the first part of the survey, many e-gamers treat games as the source of prestige, or as "pure" individual entertainment (only one third of e-gamers play multiplayer games). Thus, in general (56.15% of respondents), e-gamers do not contact each other with regard to matters which are not related to playing games (Fig. 3). The possible examination of the reasons for the increase of prestige due to playing computer games is worth taking into consideration in the future. Considering the large variety of social networking, the level of responses indicating the fact of forming circles

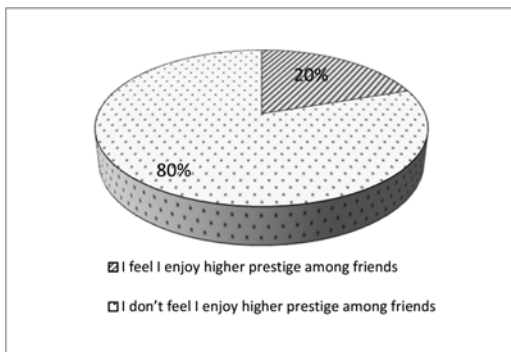


Figure 1. Playing and higher prestige among friends and acquaintances.
Source: The authors' own work (n = 254).

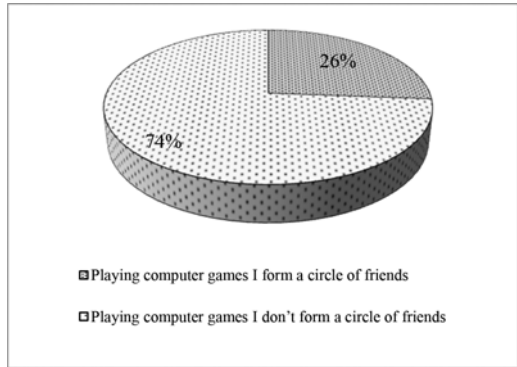


Figure 2. Forming circles of friends around playing computer games.
Source: The authors' own work (n = 254).

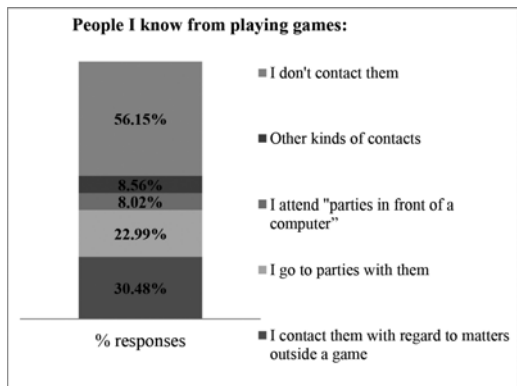


Figure 3. Social behavior towards other players.
Source: The authors' own work, (n = 254), multiple answers were possible.

of friends based around gaming is relatively low. One of the reasons for it may be strong identification of gamers with the virtual world generated in the game, and not with the real world which surrounds them.

However, on the other hand, more than 30% of respondents contact their friends also outside games. It amounts to 4 percentage points more than in the case of the circle of friends formed around the games the respondents play, which is a positive phenomenon, typical for this kind of entertainment. E-gamers indicated the fact that over 22% of them go to parties with the group of friends they share e-gaming interest with. Based on the above indication we may conclude that for e-gamers other e-gamers are attractive partners to maintain contact with, both in terms of virtual and direct contact. Slightly over 8% of gamers admit that they attend LAN parties ("parties in front of

a computer”), which is a specific way of spending time with other e-gamers, consisting in chatting and usually drinking hard drinks. Similarly—around 8% of respondents declare that they spend their time meeting friends, being involved in other kinds of activities, e.g. gaming conventions.

Subsequent questions concerned the e-gamer’s emotions after playing a game in two situations: if a gamer won or if a gamer lost (Fig. 4). The situation when the e-gamer wins a game affects his mood: 62.03% of respondents believe that they feel better after they play the game than before starting the game, and 36.36% of interviewees believe they feel the same. In the case of a game they lose, almost 38% of e-gamers feel worse than before starting the game, and over 57% say that they are indifferent to the situation. Low—which is understandable—are the indications of good mood after losing a game (4.81%) or bad after winning the game (1.60%). It is quite natural considering the wide popularity of the casual games where winning a game is a relatively simple task, and the e-gamer needs a few minutes of practice to start playing the game. In total, the obtained findings indicate a deep involvement of e-gamers in playing the games, even though the ratio of over 50% “indifference” in the case of losing a game may be worrying.

Interesting results have been obtained in the case of responses to the queries concerning the influence of factors related to studying or working, such as: interpersonal, learning and memorizing or management skills. Despite the fluctuating levels of negative responses (52–68%) in this case (they did not change as a result of participating in games), the increasingly larger number of e-gamers indicate also positive effects of games on their

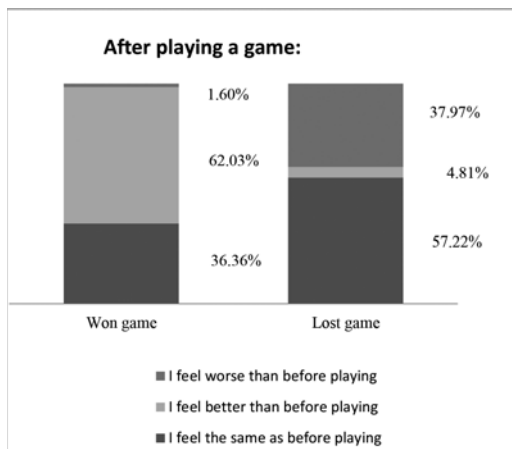


Figure 4. E-gamers’ mood after playing a game. Source: The authors’ own work (n = 254).

skills. The tendencies indicated by respondents are as follows:

- 28% improvement of interpersonal skills,
- 45% improvement of learning and memorizing skills,
- 36% improvement of management skills.

The participation of neutral skills (the skills which have not changed) is marginal (ref.: Fig. 5). The score is close to the one recorded in a similar extended nationwide study carried out in recent years (Chmielarz W. & Szumski O. 2016). The idea of the in-depth research into the reasons of the perceived improvement of skills, as indicated by the gamers is worth considering. The next element is an assessment of the degree of subjectivity of such an impression.

In order to further analyze the problem, the respondents were also asked about the influence of games on their selected psychophysical skills such as: divisibility of attention, reflex, the speed and the accuracy of decision-making, courage and stress resistance. According to the respondents, the greatest changes were indicated with regard to reflex – 59.89%. The scores above 50% were reached in the case of opinions on the positive influence of games on the speed of decision-making (56%) and divisibility of attention (55%). The smallest group of people (21%) marked that they felt more courageous as a result of playing computer games. The same number indicated the fact that, in their opinion, nothing has changed due to playing computer games. While an absolute majority of e-gamers

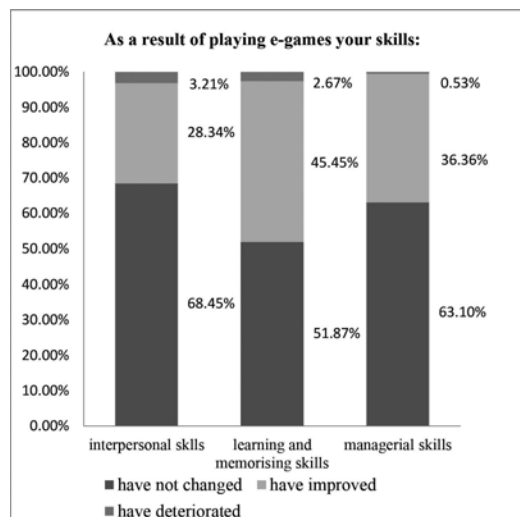


Figure 5. The effect of playing games on the e-gamers’ skills. Source: The authors’ own work (n = 254).

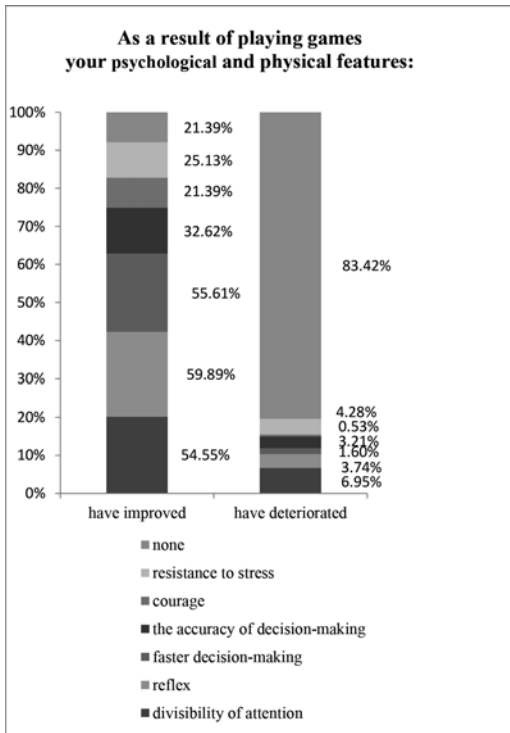


Figure 6. The influence of playing computer games on e-gamers' skills.
Source: The authors' own work, (n = 254), multiple answers were possible.

(83%) claim that playing computer games does not cause any deterioration of their psychological and physical qualities. Among the remaining responses the greatest number indicated deterioration of divisibility of attention (6.95%) or stress resistance (4.28%). The deterioration of the remaining psychophysical features was of marginal importance. In the case of the speed of decision making and multitasking they may be relatively easily verified. Another issue is the fact whether the positive influence is indicated also outside the game.

The question concerning the experience of discovering other people's cheating was of a slightly different character. The results are illustrated in Fig. 7. Nearly half of e-gamers (47.59%) provided negative responses to the query; however, it should be noted that some of the e-players may have not been aware of this phenomenon, especially in the case of games with high-speed game-play, among others FPS. The remaining respondents noticed some of the most obvious cases of cheating. They most frequently pointed to cases of fraud such as: map hack (37.43%), the function of flying (36.36%) as well as removing limitations or disadvantages in

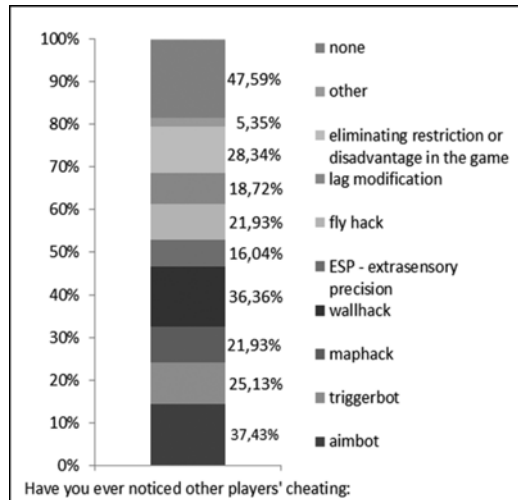


Figure 7. Cheating in computer games.
Source: The authors' own work, (n = 254), multiple answers were possible.

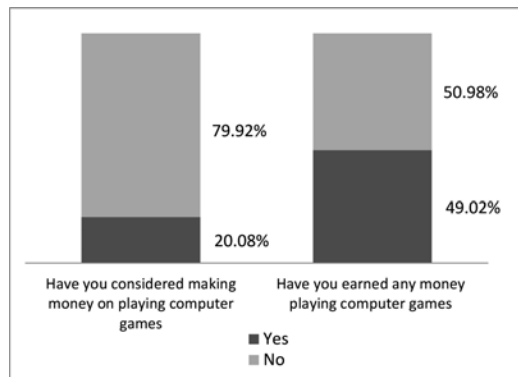


Figure 8. Earning money on participating in computer games.
Source: The authors' own work (n = 254).

a game (28.34%). Among the functions which were not included in the survey, the respondents indicated e.g.: money hack, teleport hack which was not foreseen in the game, etc.

The authors also examined the economic aspects of participating in computer games in terms of the willingness to make money or earn real money by participating in computer games (Fig. 8). We may notice a specific logical contradiction present in the findings. Namely, nearly 80% of e-gamers claim that they have not considered the possibility to earn money playing computer games, and simultaneously only 51% of the same group of e-gamers admit that they have not earned any money playing

computer games. Thus, the responses would indicate that some e-gamers were earning money without being aware of such an opportunity (?!). The second possibility is that it is merely a matter of declaration that the respondents have not considered earning money while playing computer games, and still they were doing it. An important element of the future in-depth studies should be the possible ways of making money on playing games, i.e. whether they are mainly limited to selling virtual items or we are presented with a broader spectrum of possible options.

Few people among the e-gamers know the names or pseudonyms of professional e-gamers (ref: Fig. 9). More than 82% claim that they do not know the most popular idols among players. Among the remaining group of respondents, the greatest number of them (13.78%) have heard about Brian Lewis (Astro) and Johnathan Wendel—Fatal1ty - (7.48%). The survey participants also indicated the names of other popular e-gamers not included in the list. The obtained results confirm the earlier observations concerning switching to simple entertainment games, as a result of marked changes in e-games technology.

Further problems which the survey participants had to comment on concerned the e-gamers' active and passive participation in Polish and international events related to e-games industry (meetings, trade fairs/shows, games, championships, etc.). The obtained findings turn out to be surprising. Among students of the Faculty of Management in the University of Warsaw and the students of B.Sc.

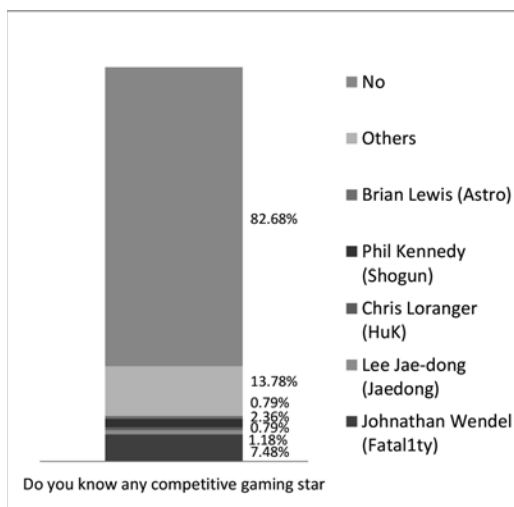


Figure 9. The level of e-gamers' knowledge concerning the most popular professional players of computer games. Source: The authors' own work, (n = 254), multiple answers were possible.

studies in the Vistula University around 85–98% of respondents never participate in such events! The highest score (nearly 8%) was indicated in the case of single or multiple times (almost 7%) when the respondents passively participated in the events connected with e-gaming in Poland. None of the members of the examined group admitted to regular, active participation in the events in Poland: 3% of respondents participated actively in the events a few times and 4% only once. The situation concerning international events is even worse: less than 3% participated in such events several times, and around 2% participated in such an event once; an active and repeated participation was indicated only by 0.39%, and 1% of respondents claim that they participated in such an event once. There remains one more problem to examine—what kind of environment do e-gamers who regularly participate in such events come from? And simultaneously there appears a chance to take advantage of the circumstances: a huge market gap which exists in the area creates the opportunity which should not be wasted.

The last group of survey questions concerned the e-gamers interests connected with playing games and other pastimes. In the first case the questions concerned additional hobbies connected with games they participated in, for example: writing stories about games, recording and publishing game videos, mixing gaming music or video clips from the games. And here e-gamers rather

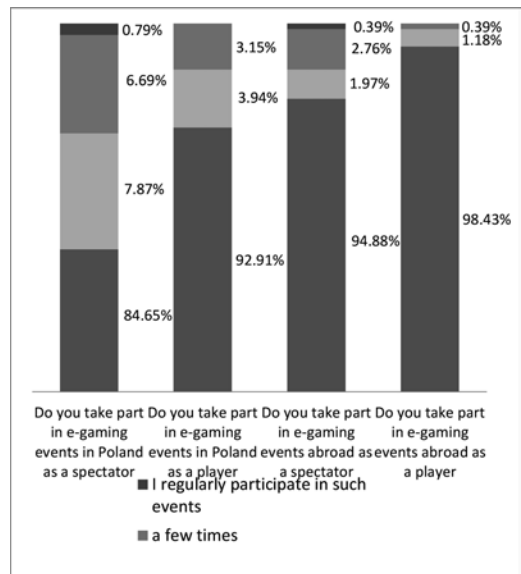


Figure 10. Passive and active participation in national and international events connected with e-gaming. Source: The authors' own work (n = 254).

concentrate on the game itself, and not on related activities, connected with the games. Almost 94% of e-gamers are not interested in it at all. The marginal number—close to 4%—record and publish game videos, mainly in the Internet. This fits within the trend known from the Web 2.0 concept, according to which a few percent of people create content for the remaining ninety percent. The findings are presented in Fig. 11.

A slightly higher level of interest is indicated in the case of passive reception of the content related to fan fiction (e.g. zins, game videos, game-related amateur films). 75% of e-gamers claim that they have never been interested in it, but 15.75% of respondents are interested in it at present, and over 9% are recipients of such content (Fig. 12). The passive reception of the content is usually connected with the cult games such as e.g. The Legend of Zelda series. Fan fiction presents itself as a niche phenomenon, and thus, we may conclude

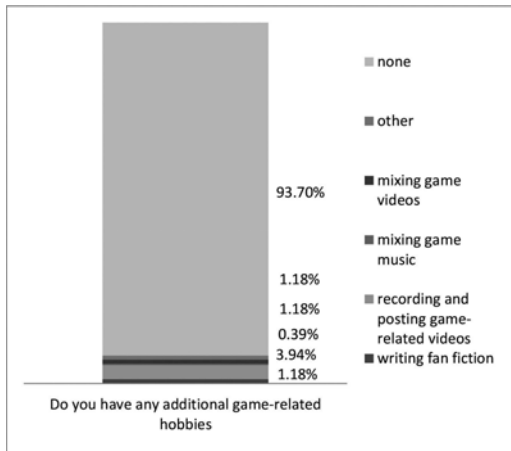


Figure 11. Additional hobbies connected with games. Source: The authors' own work (n = 254).

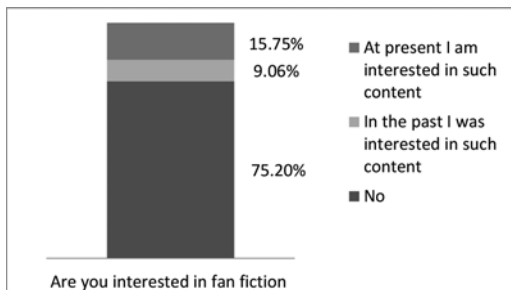


Figure 12. The reception of fan fiction content among the respondents. Source: The authors' own work (n = 254).

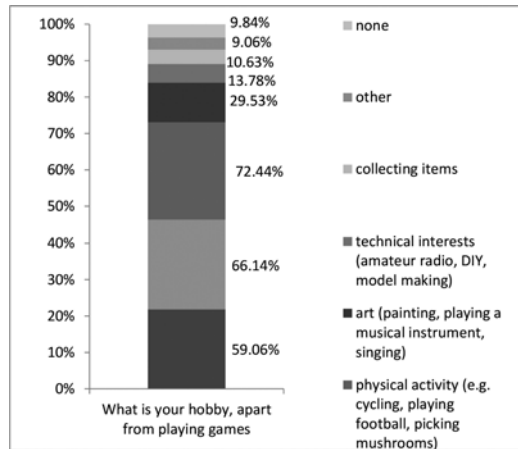


Figure 13. Additional hobbies of e-gamers. Source: The authors' own work, (n = 254), multiple responses were possible.

that 15.75% of respondents indicating their interest in such content is a relatively high score.

If we consider hobbies which go beyond the participation in computer games, it turns out that the e-gamers have very broad interests. The greatest number of people (72%–85% in other studies) are keen on physical activity (sport, leisure). Over 66% watch films; reading books as a hobby was indicated by 60% of respondents (the score is 10 percentage points lower than in the case of other studies e.g. *jestemgraczem*). Over 29% of survey participants are actively engaged in hobbies related to arts, such as e.g. painting, playing a musical instrument, etc. and almost 14% of them are interested in technical hobbies (such as model building or DIY, etc.). Nearly 11% pursue collection hobbies. Less than 10% of e-gamers admitted that they do not have any other hobbies apart from playing computer games.

4 CONCLUSIONS

The conducted research and the presented findings point to the following conclusions:

- almost all respondents (over 99% of the sample) in the current study were students, which was reflected in the obtained scores. The older the students, the weaker interest in completing the questionnaire or its findings. It is caused by the increasing number of tasks connected with studies as well as the heavy workload connected with regular or temporary work (nearly 36% of working students). The latter is confirmed in the scores of other surveys (*#JestemGraczem* 2014,

Marketing przy Kawie 2014, Newzoo 2014, Żywiczyńska E. 2014b), despite the fact that, in total, fewer than 25–16% students participated in the study (even though it was always the largest group of players),

- among people who completed questionnaires there were markedly more women (almost 60%) than in other survey studies (around 43–48%) (Żywiczyńska E. 2014a), conducted two or three years ago. Thus, we may conclude that there occurs a specific change with regard to the number of women playing computer games. Naturally, we should also be aware of the fact that the present study examined mainly the responses of students of economic faculties, and in this case the general number of female students in these faculties is greater than men. Still, the survey included also the option I don't play computer games, which the women could indicate,
- participation in computer games does not significantly raise an e-gamer's prestige; also, it does not influence the creation of circles of friends among e-gamers, and over a half of e-gamers admit that they do not contact each other outside the game,
- after winning a game the positive emotions of e-gamers are significantly on the increase; losing the game, on the other hand, does not result in creating strong emotions, 50% percent of the survey participants claim that they remain indifferent to the outcome of the game,
- e-gamers positively evaluate participation in games: almost half of respondents feel that their learning and memorizing skills have improved,
- in the opinions of respondents, games have improved mainly such psychophysical skills as: reflex, the pace of decision-making and divisibility of attention. Over 80% of people maintain that playing games does not deteriorate their psycho-physical qualities,
- over half of e-gamers have noticed the cases of cheating when playing games and admit that various hacks may be applied by players,
- computer games to a greater and greater degree are, even though e-gamers are trying to prove otherwise, seen as the source of income (e.g. being involved as a player, or selling a game account, etc.),
- in total, e-gamers generally do not know the idols of e-sports circles (more than 80% of respondents); they do not take part, both as players or spectators, in the e-sports-related events in Poland and abroad (over 90%); also, they do not have any additional hobbies (such as: writing fan fiction, recording and posting game videos, mixing gaming music or mixing game video clips)

connected with this kind of entertainment (this indication was reported in more than 90% of cases),

- relatively more people (around 25%) are interested in the fan fiction content due to the fact that they do not require such intense player's engagement as the above said activities,
- the surveyed e-gamers have broad interests, outside the area of gaming (they are mainly active with regard to physical exercise (sport, leisure), they watch films, read books or they are interested in art).

The findings of the second stage are not so positive as those obtained in the preliminary stage of the research. They tend to support the thesis concerning the fact that games are treated as a form of entertainment rather than sport, and they belong to a domain of amateurs, rather than people who would like to take up playing games professionally. In order to move this form of entertainment to the next level, we need to examine the reasons why the group comprising enthusiasts and people who are serious about e-games is so small, and the further studies should include fan clubs and groups of players participating in the tournament games (both home and abroad).

REFERENCES

- [1] #JestemGraczem 2014. Badanie graczy w Polsce, <http://www.jestemgraczem.com/wyniki>, access, January 2016.
- [2] Akcjonariat Obywatelski 2013. Polski rynek gier komputerowych na tle rynku światowego, <http://akcjonariatobywatelski.pl/pl/centrum-edukacyjne/gospodarka/1033.Polski-rynek-gier-komputerowych-na-tle-rynku-swiatowego.html>, access, January 2016.
- [3] Chmielarz W. & Szumski O. 2016. Charakterystyka e-graczy i ich preferencji w grach komputerowych, accepted for publication as conference materials: Innowacje w zarządzaniu i inżynierii produkcji "2016", Zakopane, 2016.
- [4] Chmielarz, W. 2015a. Porównanie wykorzystania sklepów internetowych z aplikacjami mobilnymi w Polsce z punktu widzenia klienta indywidualnego (Comparison of the Use of Mobile Applications Websites in Poland from the Point of View of Individual Client) in: Innowacje w zarządzaniu i inżynierii produkcji edited by R. Knosala, in: Vol. II, Part IX Inżynieria ja-kości produkcji i usług, Oficyna Wydawnicza Polskiego Towarzystwa Zarządzania Produkcją, Opole, 2015, pp. 234–245.
- [5] Chmielarz, W. 2015b. Study of Smartphones Usage from the Customer's Point of View, *Procedia Computer Science*, Elsevier, Vol. 65, 2015, pp. 1085–1094.
- [6] GRY-OnLine 2004. Klasyfikacja gier, <http://www.gry-online.pl/S018.asp?ID=208&STR=2>, access, January 2016.

- [7] GRY-OnLine 2014. 13,4 miliona graczy w Polsce i inne informacje o naszym rynku, <http://www.gry-online.pl/S013.asp?ID=82806>; access, January 2016.
- [8] IT-Pomoc.pl 2016. Czym jest gra kompute-rowa, <http://it-pomoc.pl/komputer/gra-komputerowa>; access, January 2016.
- [9] KIPA 2016. Definicje gier komputerowych, <http://www.kipa.pl/index.php/promocja-filmu/gry-komputerowe/definicje-gier-komputerowych>, access, January 2016.
- [10] Marketing przy Kawie 2014. Jacy są Polacy grający w gry komputerowe? <http://www.marketing-news.pl/message.php?art=43734>, access, January 2016.
- [11] Mijal M. & Szumski O. 2013. Zastosowania gier FPS w organizacji, in: Chmielarz W., Ki-sielnicki J., Parys T. eds), *Informatyka @ przyszłości*, Wydawnictwo Naukowe WZ UW, Warsaw 2013, pp. 165–176.
- [12] Newzoo 2014, *Global Games Market Report*, <http://www.newzoo.com/product/global-games-market-report-premium/>, access, January 2016.
- [13] PTBG 2010. *Homo Ludens* 1/(2), <http://ptbg.org.pl/HomoLudens/vol/2/>, access, January 2016.
- [14] Wiedza i Edukacja 2009. *Analiza gier*, <http://wiedzaiedukacja.eu/archives/tag/analiza-gier>, access, January 2016.
- [15] Wikipedia 2016. *Gra komputerowa*, https://pl.wikipedia.org/wiki/Gra_komputerowa, access, January 2016.
- [16] Zajac J. 2014a. *Jestem graczem w social media*, at: <http://blog.sotrender.com/pl/2014/12/jestem-graczem-w-social-media/>, access, January 2016.
- [17] Żywiczyńska E. 2014a. *Co tak naprawdę wie-my o graczach*, at <http://zgranarodzina.edu.pl/2014/10/12/co-tak-naprawde-wiemy-o-graczach/>, access, January 2016.
- [18] Żywiczyńska E. 2014b. *Optymizm czy myślenie ży-czeniowe. Zaskakujące wyniki badania #jestemgra-czem*, at: <http://zgranarodzina.edu.pl/2014/12/20/optymizm-czy-myslenie-zyczeniowe-zaskakujace-wyniki-badania-jestemgraczem/>, access, January 2016.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

The utilization of the HR analytics by the high and mid-level managers: Case from Eastern Poland

Monika Wawer

John Paul II Catholic University of Lublin, Lublin, Poland

Piotr Muryjas

Lublin University of Technology, Lublin, Poland

ABSTRACT: Measuring and analysing various business processes and people activities, are the ones of the key factors having a critical influence on the results achieved by today's modern firms. Due to the growing amount of data, its complexity and variety, the managers have to apply the new Information Technology (IT) tools and methods of data analysis which enable one to find the relationships between employees' performance and the organisation's outcomes more effectively. The high and mid-level managers should be strongly engaged in these activities. They are directly responsible for results in many areas of the Human Resource (HR), but the main problem is that they do not understand the immense potential which the analytical thinking offers.

The aim of this article is to analyse the utilization of the HR analytics by high and mid-level managers in the contemporary organisations. The findings of research confirm that almost half of surveyed organisations do not apply the analytical approach in the HR management. The results also show differentiated approach of high and mid-level managers to the utilization of the analytics in the HR areas.

1 INTRODUCTION

“If you can measure something, you can manage it. If you can manage something, you can achieve objectives”—this sentence is as true today as ever before. The necessity to measure and analyse different processes and the people who are involved is a key factor that has a critical influence on the organisation's outcomes.

The analytics in today's modern organisations play a crucial role in the management processes. Many business functions such as production, supply, sales, marketing, and customer relationships management apply analytical approach utilizing various metrics, Key Performance Indicators (KPI), score cards and Information Technology (IT) tools to improve the general performance (Parmenter 2010, Kaplan & Norton 2005). These areas are often perceived as traditional places where the decision-making process is supported by the results of analytics. Nevertheless, today the analytical approach should be broadly applied in all areas that influence the organisations outcome. One of them is Human Resource Management (HRM) which seems to fall behind the above-mentioned areas.

In many organisations the HR function has traditionally been seen as a cost centre rather

than a value creating source (Pemmaraju 2007). Nowadays it is necessary to shift this point of view and treat the human resources as a valuable strategic partner that helps the company achieve its goals (Philips & Philips 2008). The key to success is to identify and measure the HR “deliverables” that support corporate strategy—and the HR systems that create those deliverables (Becker et al. 2001). That means that contemporary managers have to apply the new Information Technology (IT) tools and methods of data analysis which enable one to find the relationships between people and the organisation's outcomes more effectively.

The exceptional importance in this process has the high and middle-level managerial staff (Song et al. 2014). They are directly responsible for the results of the work of their subordinates. Their main task is to realise the most important functions of human resource management. According to Armstrong (2006), the key HRM areas are: human resource planning, recruitment and selection, introduction to the organisation, formulating and implementing learning and development strategies, performance appraisal and performance management, talent management and career management, motivation and compensation, reward management, employee benefits, pensions and allowances, release from the organisation.

The role of the modern high and middle-level managers in each of the areas mentioned above is fundamental. The more often they use HR analytics, the more effective they will acquire, motivate, develop and retain employees (Kapoor & Sherif 2012). By understanding the past performance, current results, and future possibilities, analytics-driven managers can help companies achieve better business outcomes.

The aim of this article is to evaluate the level of utilization of the human resource analytics by high and middle-level managers in the contemporary organisations.

2 BUSINESS ANALYTICS IN TODAY'S ORGANISATIONS

A digital universe—is probably the best description of today's world. Data is generated constantly and everywhere by people, processes, and devices. According to a study conducted by International Data Corporation (IDC), the amount of data will increase 50-fold from the beginning of 2010 to the end of 2020 (Gantz & Reinsel 2012). But possessing data does not create its value. Only information, that we can discover using this data, has the real value. IDC estimates that by 2020, as much as 33% (8% more than today) of the digital universe will contain information that might be valuable if analysed.

Along with the growth in the volume of data, its complexity and variety have risen as well. Decision-making in these conditions has become a very difficult task and requires support that will allow an effective and efficient utilization of this resource. The capabilities of managers do not allow them to identify which data is significant and to discover relationships between that data as well as create the multidimensional views of the organisation. This is why in the analysis of such huge resources the specialized IT tools have to be used in order to transform data into useful and valuable business information. The active and widespread data utilization in decision-making is the strong evidence of applying the data-driven management. One of its most significant elements is Business Analytics (BA).

The business analytics is defined as a set of methods that transform raw data into action by generating insights for organisational decision making (Liberatore & Luo 2010). Watson (2009) enhanced this definition about the technical aspects and claims that BA is a broad category of applications, technologies, and processes for gathering, storing accessing, and analysing data to help business users make better decisions.

Deloitte Corporation confirms the growing interest in analytics and its critical importance in

decision-making processes. 84% of respondents expressed the opinion that the utilization of analytics has increased the competitiveness of their organisations, 25% of them stated that the growth is very high and 30% gauged it as high (Davenport 2013). Moreover, the analytical approach and BA in managing the contemporary organisation have been indicated by 77% of CIO (Chief Information Officer) in newest Deloitte's survey as the most important area of technology that will have a significant impact on business in the next two years (Kark et al. 2015). There has never been a better time to understand the importance of analytics in the contemporary business.

From historical point of view, the analytics were primarily utilized to create the quantitative description of the past and to answer the following questions: what happened, how often and why? (Fitz-enz 2009). This type of analytics is mainly supported by data stored in the Enterprise Resource Planning (ERP) systems that have dedicated modules to carry out the business processes (Parry 2011). Along with the shrinkage of the decision window and the IT technology development, the analytics evolved to deliver the information about the present and predict the future (Fitz-enz 2010, 2014, Corne et al. 2012). Depending on the maturity of the organisation, the managers can analyse business facts using techniques with various levels of advancement (Fig. 1).

Creation of the deep and full insight requires the consideration of a broad range of data and its sources. The ability to generate information from both internal and external data is crucial in today's world (Drucker 1995, March & Hevner 2007). Although much has already been achieved through the use of analytics in the contemporary organisations but there is still much that can be achieved with existing data collected in the organisation's IT systems (Angrave et al. 2016).

The utilization of analytics in many business areas (mentioned in the introduction) gives much evidence that it is the right direction of activities that lead to performance improvement on the whole organisation level as well as on the level of

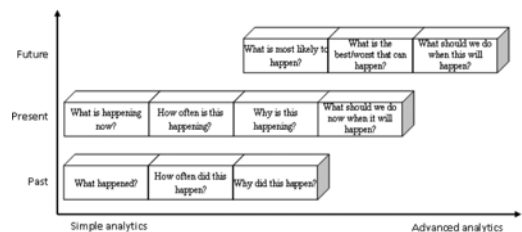


Figure 1. Dependence between the level of advancement of analytics and the time horizon of their use.

particular departments. The key question is how analytics can help to make more effective the HR management and how to increase the HR contribution to the success of the firm?

3 HR ANALYTICS APPROACH—FAD OR NECESSITY?

The analysis of the subject's literature (e.g. SAP 2016, Rasmussen & Ulrich 2015, Moon & Prouty 2015, Demetriou et al. 2015, Smith 2013) indicates that HR analytics are still in infancy stage and very few organisations are actively implementing people analytics capabilities to address complex business and talent needs. Angrave et al (2016) notice that many firms have begun to engage with HR data and analytics, but most of them have not progressed beyond operational reporting. According to Moon from Aberdeen Group (2015b), 44% of organisations cite that one of the main reasons of their HR function struggles with systematically using analytics is a lack of people who understand how to interpret analytics and how to turn them into actionable insights. There are few examples of more advanced utilizations of analytics (predictive and/or prescriptive analytics) in HR management (e.g. Fitz-enz 2010, 2014, Kapoor & Kabra 2014).

Some authors (e.g. Rasmussen & Ulrich 2015) note that the HR analytics in the current form has a risk of being a fad that fades. The main pitfalls mentioned by them, which testify to it, are: a lack of analytics about analytics, mean/end inversion or data fetish, academic mindset in business setting, running HR analytics only from HR department and a journalistic approach to HR analytics. Boudreau (2014) claims that the implementation of HR metrics and analytics systems is one of the arenas reflecting HR's role in shaping strategy and building effective HR skills, where changes are the slowest.

However, the literature of the subject (Aral et al. 2012, Levenson 2011) and many surveys conducted by leading companies such as Aberdeen Group, Deloitte Consulting, PwC, KPMG and McKinsey & Company deliver the evidences that the awareness of importance of analytics is growing in the organisations and the need of analytics in HR area is much stronger today than in the past. According to Deloitte, 75% of surveyed companies believe that using people analytics is important (Demetriou 2015). Aberdeen Group states that nearly 80% of all respondents indicated that having analytics about their workforce was critical to their organisation's business strategy (Moon 2015a). KPMG informs that 56% of HR functions report an increase in using data analytics compared to three years ago, 31% plan to implement new technology to support this and for 23% of the

respondents the adoption of data analytics would be their main focus in the next three years (KPMG 2013). Moreover, 65% of organisations have applied advanced analytics to improve efficiency of the HR functions and 70% expect to begin using or increase their use of advanced analytics to inform HR decisions in the next three years. The proportion of respondents who say their organisation's HR function "excels" at providing insightful and predictive analytics increased from 15% in 2012 to 23% in 2014. Over the same period, the percentage the ones who say the function excels at measurably proving the value of HR to the business has increased from 17% to 25% (KPMG 2015).

Another evidence of rising interest in HR analytics delivers PwC. 86% of PwC Saratoga participants reported that creating or maturing their people analytics function is a strategic priority over the next one-to-three years. And nearly one-half (46%) of those organisations already have a dedicated people analytics function (PwC 2015). In turn McKinsey reports that HR analytics is one of the top ten critical future priorities in the HR area in the next 2–3 years. Among these respondents who represent this point of view, 36% of them are already doing this and another 36% have plans with a high priority to do analytics (McKinsey 2012).

The Chartered Institute for Personnel and Development emphasizes that analytics is a 'must have' capability for HR managers, that creates value from people and a pathway to broadening the strategic influence of the HR function (CIPD 2013). Moreover, Ulrich & Dulebohn (2015) mention that creating HR analytics focused on the right issues and gaining the skills to comprehend how to use metrics to support decision-making are important domains for HR investments. This point of view is also shared by Cohen (2015) who emphasizes the key role of HR analytics in the contemporary business acumen that influences the ability to understand and to apply information to contribute to the organisation's strategic plan.

HR analytics is strongly supported by Business Intelligence (BI) systems, which can be seen as the technological foundation to conduct business analytics (Chiang et al. 2012, Lim et al. 2013). New information technology tools and methods of data analysis enable HR professionals to find the relationships between people and organisation's outcomes more effectively (Muryjas & Wawer 2013). Advances in technology are creating opportunities for managers to start a new kind of dialogue about the link between the people and their performance (Gardner et al. 2011). This dialogue allows one to determine the impact of HR activities on the aim achievement and it will create business value for the enterprise. The relationship between BA and HR management has been mentioned by Laursen &

Thorlund (2010). They indicate 4 scenarios of BA use in the organisation:

- BA and HR management separation—BA does not deliver data to the strategic level, it is only used to answer some questions on the operational level,
- Passive support of the HR management by BA—the only role of BA is to produce reports to support the strategy performance,
- Dialogue between BA and HR management—the results of BA may modify the management activities,
- Interpenetration of BA and HR management—results of BA are treated as a crucial resource of the organisation, which determines the HR management.

The adoption of the certain scenario depends on the maturity of the organisation to use BA in defining and realisation of the HR management strategy (Fig. 2).

The importance of BA in HR management has been emphasized by Fitz-enz, who says that “the human capital analysis and predictive measurement can provide this information and are, therefore, critical for business success in this global marketplace” (2010).

There are many arguments for implementing HR analytics in the modern enterprise but the main problem is that HR professionals do not understand the potential which the analytical thinking offers. One of the reasons is that there is little evidence for successes of HR analytics implementation. However, to be a successful catalyst for change, HR should not only be capable of analysing and interpreting human capital metrics and analytics, but to also be capable of recommending and implementing interventions to drive organisational effectiveness (Moon 2015b).

Based on the previous discussion, it is possible to state that the achievement of organisation’s aims with the utilization of HR analytics requires an engagement of managers at different levels. According to Gary & Wood (2011), every manager has knowledge structures that impact the perception, information processing, problem solving, and

decision making, influencing the organisational learning capability and firm performance. It is also worth to emphasize that high-level managers should share the expert knowledge with the mid-level managers and propagate the analytical approach among them to increase their analytical competences to utilize the analytics in HR management. However, in many organisations middle managers do not receive the right information or do not take appropriate actions aimed at the increase of the efficiency of HR management.

The aim of our research is to answer the question whether the high and mid-level managers are engaged in the same way in the utilization of analytics in the HR management. The following main hypothesis has been defined:

H: There is a dependence between utilization of HR analytics in the organisation and the level of a managerial position.

In order to verify this hypothesis, three detailed hypotheses have been formulated:

H1: There is a dependence between the frequency of using the analytics in different HR areas and the level of a managerial position.

H2: There is a dependence between the types of analytics used in different HR areas and the level of a managerial position.

H3: There is a dependence between the appraisal of the level of benefits of HR analytics utilization in the enterprise management and the level of a managerial position.

In order to verify these detailed hypotheses, managers have been asked the following questions:

1. How often do you utilize the analytics in different HR areas? (Possible answers were: very often, often, rarely, never).
2. Which types of analytics do you use in particular HR areas? (Possible answers were: descriptive analytics, predictive analytics, both of them, none of them).
3. In your opinion, what is the level of benefits of utilization of HR analytics in the enterprise management? (Possible answers were: high, middle, low, none).

4 RESEARCH METHODOLOGY

In November 2015, we sent e-mails with a questionnaire to 237 potential organisations that reside in Eastern Poland. The survey was conducted to January 20, 2016 on the territory of three provinces, namely Lubelskie, Podlaskie, and Subcarpathian. We received a total of 73 responses to this study. Every organisation was represented by one respondent. To ensure the quality of data, we carefully

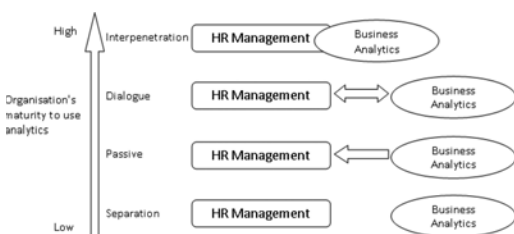


Figure 2. Scenarios of BA and HR management in dependence on the organisation’s maturity.

scrutinized and verified all respondent's entries to ensure that the study includes only fully completed questionnaires. We excluded those which did not contain answers to every question and accepted 61 questionnaires to the further analysis.

The surveyed managers represented organisations in the financial and insurance services (16.39%), consulting (8.20%), technology/telecommunication (6.56%), entertainment (4.92%), health care (9.84%), education (8.20%), consumer products/retail (21.30%), manufacturing (18.03%), and public administration (6.56%) sectors.

Most of the survey participants were mid-level executives (75.41%), and 24.59% represented high-level managers. The detailed description of the respondents is presented in Table 1.

Table 1. Respondents' description.

Category	Total	High level	Middle level
Gender			
Woman	47.54%	46.67%	47.83%
Male	52.46%	53.33%	52.17%
Age			
20–35 years	34.43%	20.00%	39.13%
36–45 years	39.34%	20.00%	45.65%
46–55 years	26.23%	60.00%	15.22%
Seniority			
till 3 years	31.15%	20.00%	34.78%
4–8 years	34.43%	13.33%	41.30%
9–13 years	22.95%	26.67%	21.74%
14–18 years	3.28%	6.67%	2.17%
19 and more years	8.20%	33.33%	0.00%
Organisation size			
10–49 persons	11.48%	0.00%	15.22%
250–500 persons	16.39%	0.00%	21.74%
50–249 persons	45.90%	46.67%	45.65%
more than 500 persons	26.23%	53.33%	17.39%
Core business			
Production	19.67%	20.00%	19.57%
Services	54.10%	46.67%	56.52%
Trade	14.75%	13.33%	15.22%
Other	11.48%	20.00%	8.70%
Place of organisation			
City to 10 thousand citizens	4.92%	0.00%	6.52%
City 10–50 thousand citizens	9.84%	0.00%	13.04%
City 50–250 thousand citizens	13.11%	0.00%	17.39%
City more than 250 thousand citizens	72.13%	100.00%	63.04%
Property			
Private with foreign capital	39.34%	40.00%	39.13%
Private with polish capital	40.98%	46.67%	39.13%
State treasury	18.03%	13.33%	19.57%
Community	1.64%	0.00%	2.17%

The statistical analysis of the survey results was performed using the R-3.1.2 environment. The dependence between categorical variables was examined using Pearson's chi-square test for independence with Yates' correction for discontinuity. The accepted statistical significance level is $p < 0.05$.

5 FINDINGS AND DISCUSSION

5.1 Frequency of utilization of analytics in different HR areas depending on the level of management position

Conducted survey confirms that HR analytics are not yet commonly utilized in the contemporary organisations. Among all the surveyed firms, taking into consideration the total results, only 17% use the analytics very often, and 29% often. Unfortunately, 34% said they utilize HR analytics rarely and 20% never apply them. Such distribution confirms that HR analytics are in the area of interest of managers only in few firms nowadays.

Considering the aim of this survey it is important to analyse the utilization of HR analytics on different levels of management. The key question is whether the mid-level and high-level managers utilize the HR analytics with the same frequency? The results of this survey related to this question are presented in the Table 2.

The analysis of these results allows us to state that 28% of the high-level and 14% of the mid-level managers utilize HR analytics very often. This proportion may indicate that the decision making on the high-level management requires the support more often than the mid-level. However, other data representing the overall view on the HR area does not confirm the statistically significant dependence between the frequency of utilization of the analytics and the level of management positions.

The analysis, that takes into consideration ten main HR areas, delivers very interesting results. Figure 3 shows how often the high and mid-level managers apply the analytics in the following areas:

1. human resource planning (A1),
2. recruitment and selection (A2),
3. introduction to the organisation (A3),
4. formulating and implementing learning and development strategies (A4),

Table 2. Frequency of HR analytics utilization depending on the management level.

	Very often	Often	Rarely	Never
Middle level	14%	31%	34%	21%
High level	28%	22%	31%	19%

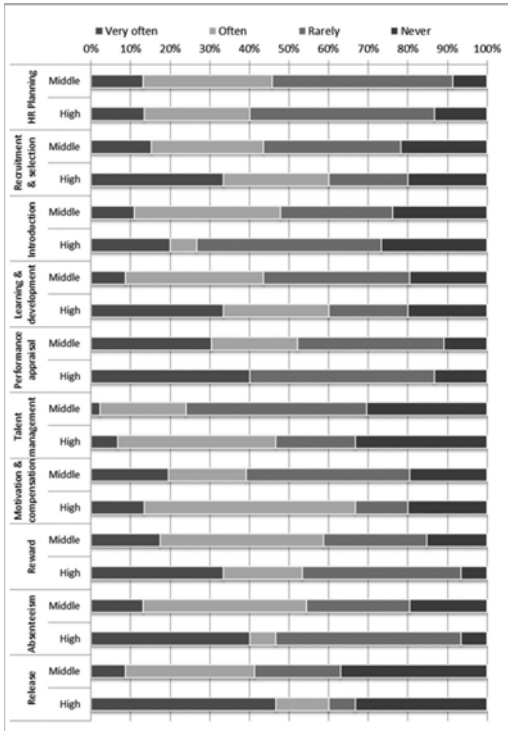


Figure 3. Frequency of utilization of analytics within each HR area depending on the level of the managerial position.

5. performance appraisal and performance management (A5),
6. talent management and career management (A6),
7. motivation and compensation (A7),
8. reward management (A8),
9. employees absenteeism (A9),
10. release from the organisation (A10).

The chi-square test does not demonstrate the statistically significant differences in answers concerning areas A1 to A6 and A8. It means the structure of the answers to the question about the frequency of utilization of analytics in these HR areas does not depend on the level of managerial position.

However, results of this test show that this type of dependence exists in case of areas A7 ($p = 0.0049^*$), A9 ($p = 0.012^*$) and A10 ($p = 0.007^{**}$).

In area A7 (Table 3) 53% of high-level managers answered “often”, while 41% of the mid-level executives said “rarely”.

For A9 area the most answers (Table 4) for high-level managers were “very often” (40%) and “rarely” (47%), while on the mid-level it was mainly the answer “often” (41%).

Table 3. Frequency of utilization of analytics in motivation and compensation area depending on the level of the managerial position.

Motivation and compensation	Total	High level	Middle level	Chi-square p
Very often	18.03%	13.33%	19.57%	$p = 0.049^*$
Often	27.87%	53.33%	19.57%	
Rarely	34.43%	13.33%	41.30%	
Never	19.67%	20.00%	19.57%	

Table 4. Frequency of utilization of analytics in absenteeism area depending on the level of the managerial position.

Employees absenteeism	Total	High level	Middle level	Chi-square p
Very often	19.67%	40.00%	13.04%	$p = 0.012^*$
Often	32.79%	6.67%	41.30%	
Rarely	31.15%	46.67%	26.09%	
Never	16.39%	6.67%	19.57%	

Table 5. Frequency of the utilization of analytics in employee release area depending on the level of the managerial position.

Release from the organisation	Total	High level	Middle level	Chi-square p
Very often	18.03%	46.67%	8.70%	$p = 0.007^{**}$
Often	27.87%	13.33%	32.61%	
Rarely	18.03%	6.67%	21.74%	
Never	36.07%	33.33%	36.96%	

In area A10 (Table 5) the most frequent answers in the group of high-level managers were “very often” (47%) and “never” (33%), while on the mid-level management—“often” (33%) and “never” (37%).

The analysis of the statistical dependencies indicates that in three HR areas the visible and significant differences in utilization of analytics by managers of high and mid-level have been observed.

The first area is motivation and compensation. Here the high-level managers, first of all, have the main impact on the organisation’s strategy and they define directions of HR shifts that shape this strategy. The mid-level executives in this area have limited privileges and therefore they are less engaged in these processes.

The second area is the employee absenteeism. The middle managers have direct contact with the employees and they are responsible for the results of their work to the greatest extent. This is probably the reason why they utilize the analytics to evaluate the absenteeism which has a strong influence on the outcomes.

The high-level managers are rather involved in the strategic activities than in the operational processes. Therefore they are aware of the capabilities and benefits of analytics and utilize them very often in the long-term analysis of the reasons of absenteeism as well as its prevention. The managers, who do not have the adequate knowledge and awareness in this area, they do not feel the need and do not perform these analyses.

The third HR area, which differentiates the approaches of managers to utilize analytics, is the employee release from the organisation. Perhaps the reason behind this fact is the same as the previously mentioned one.

In summary, due to the strategic reasons the high-level managers utilize HR analytics very often and middle managers are doing this often due to their engagement in the operational activities with subordinates. The explanation of these differences requires further research.

5.2 Types of analytics utilized in particular HR area depending on the level of management

As mentioned in section 2, analytics should be applied so as to consider two main time horizons. The first one concerns the past, which can be analysed using techniques of descriptive analytics. The second one is the future that can be foreseen by using the predictive analytics. Both types of approaches are very important in every organisation. The effective support of achievement of organisation's aims by HR management will be possible only when HR analytics will be utilized from the point of view of both time perspectives.

A conducted survey confirms that HR analytics are not properly used in the organisations. Among all the firms, taking into consideration the total results (high and mid-level managers), both types of analytics are adopted only by 8% of surveyed, whereas the descriptive analytics is performed by 28% of managers and the predictive by 35%. Unfortunately, 28% of respondents stated that they do not apply the analytics for any purpose mentioned here. This distribution of answers confirms that managerial staff does not fully utilize the capabilities of HR analytics.

From the point of view of the essence of our research, it is important to recognize the type of HR analytics which are used by the managers on different managerial levels. The key issue is why the high and mid-level managers are using the HR analytics? What is the time horizon of analytics for every manager group? The results of conducted survey are presented in the Table 6.

As results from this survey show, HR descriptive and predictive analytics are utilized only by 6% of the high-level managers and 8% of the mid-level managerial staff.

Table 6. The type of HR analytics depending on the level of managerial position.

	Descriptive	Predictive	Descriptive & predictive	None
Middle level	32%	30%	8%	30%
High level	21%	47%	6%	26%

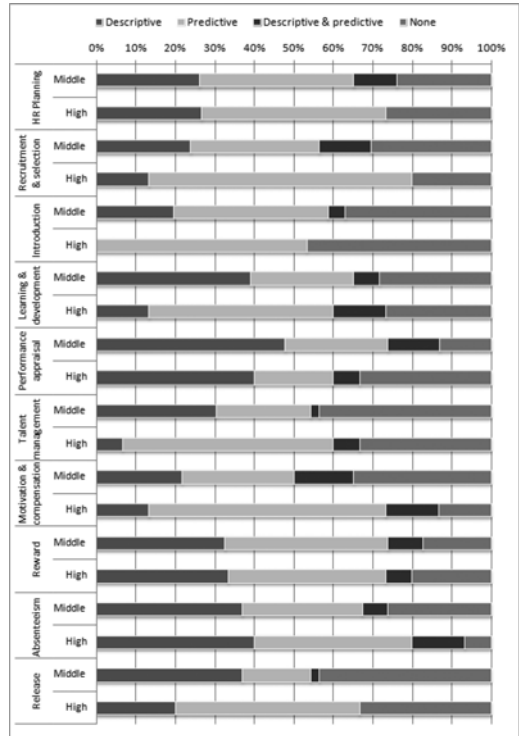


Figure 4. Types of analytics in HR areas depending on the level of management.

47% of high-level and 30% of mid-level executives declare they use the data analytics to predict the future activities in the HR areas. This difference shows that the high-level managers have a better understanding of the true meaning of the use of this type of analytics. However, the disturbing fact is that 26% of high-level and 30% of mid-level respondents never use the HR analytics in their work.

Enhancing the above reflections and taking into consideration the individual areas of HR we can observe an interesting relationship between them. Figure 4 presents the type of analytics used in particular HR areas depending on the level of managerial position.

The middle managers utilize both types of analytics most often for the analysis of motivation

and compensation (15%), performance appraisal (13%), recruitment and selection (13%), and HR planning (10%), while the high-level executives in the following areas: learning and development (15%), motivation and compensation (14%), and absenteeism (13%).

Descriptive analytics are utilized by about the same percentage of respondents in the group of middle and high-level managers in the following HR areas: performance appraisal (48% vs. 40%), reward and absenteeism (32% vs. 33%), and HR planning (26% vs. 27%). In other HR areas the middle executives use the descriptive analytics more often than high-level managers, i.e. learning and development (39% vs. 13%), release (37% vs. 20%), talent management (31% vs. 7%), recruitment and selection (23% vs. 13%), motivation and compensation (22% vs. 13%), and introduction (20% vs. 0%).

The survey results also show that predictive analytics are performed by high-level managers mainly in the following HR areas: recruitment and selection (67%), motivation and compensation (60%), and talent management (53%).

Summing-up the above results, we can state that although there is no statistically significant dependence between the types of HR analytics and the level of the managerial position, both groups of managers utilize them to analyse the activities of employees in different time horizons and in different HR areas.

5.3 *The level of benefits of HR analytics use in the enterprise management*

The third research area concerns the opinion of the managerial staff about the level of benefits of using the HR analytics in the enterprise management. The results allow us to conclude that all surveyed managers have a high awareness in this area and see the benefits of HR analytics use in the enterprise management. Half of respondents (50.82%) reported that using analytics assures the high level of benefits, followed by 40% who said that it is average. Another 8% stated that the benefits are very low.

Because of the purpose of our research and defined hypothesis it is important to compare the opinions of managers about the level of benefits taking into account their managerial position. The analysis performed with the use of chi-square test revealed that there is a statistically significant dependence ($p = 0.042^*$) between the answers of both groups of managers (Table 7).

Most high-level managers (73%) see high level of benefits which are delivered by the data analytics in management, while 13% of respondents in this group reported the average and low level

Table 7. Level of benefits of HR analytics utilization in enterprise management depending on the level of the managerial position.

Level of benefits (%)	Total	High-level managers	Mid-level managers	Chi-square p
High	50.82%	73.33%	43.48%	$p = 0.042^*$
Average	40.98%	13.33%	50.00%	
Low	8.20%	13.33%	6.52%	

respectively. The situation looks different in case of mid-level managers. Only about 43% of them stated that the benefits of use of the analytics are very high, and one-half of surveyed determine benefits as average.

6 CONCLUSIONS

In the context of presented subject literature review it is possible to state that the frequency of utilization of analytics in the HR area is still very low in many organisations. The results of our research confirm that HR analytics are not used at all by managers in 20% of surveyed firms, and 30% of respondents utilize them rarely. It means that almost a half of all surveyed organisations do not apply the analytical approach in the HR management.

The second significant conclusion from our research concerns the type of performed analytics. Less than 10% of managers utilize both types of analytics (descriptive and predictive), and unfortunately as many as 28% of the respondents do not apply any type of analytics in the management processes.

Thirdly, we should positively assess the opinions of 50% of managers who see the high level of benefits of utilization of HR analytics in the enterprise management.

These conclusions are of a general nature but are directly related to three specific research hypotheses. The conducted research indicates that the high and mid-level managers utilize analytics with various frequencies in three HR areas: motivation and compensation, employee absenteeism, and release from the organisation. This fact confirms our first hypothesis.

The survey results describing the types of HR analytics applied by managers of varying levels did not disclose the statistically significant difference. Therefore, the second hypothesis has not been confirmed. However, we can observe that high and mid-level executives use analytics differently in particular HR areas.

The last hypothesis concerning the dependence between the appraisals of the level of the benefits of

HR analytics utilization and the management level has positively been verified. The results confirm the statistically significant dependence between the answers and the managerial level.

Summing up the conclusions above, it can be stated that the main hypothesis about the various approaches of managers of different levels to the utilization of the analytics in HR areas has partially been confirmed.

While formulating these conclusions it is necessary to be aware of the occurrence of certain limitations. The analyses performed in this research paper did not take into account the specifics of the surveyed organisations that are different in size, the sector, the type of core business, and the form of ownership. Moreover, they are located in Eastern Poland that has agricultural character and is less economically developed than Western Poland.

These circumstances could affect the survey results and they should be the base for further research that might also focus on the dependences, mentioned above, in larger group of organisations located in the whole country.

REFERENCES

- Angrave, D., Charlwood, A., Kirkpatrick, I., Lawrence, M. & Stuart, M. 2016. HR and analytics: why HR is set to fail the big data challenge. *Human Resource Management Journal* 26(1): 1–11.
- Aral, S., Brynjolfsson, E., & Wu, L. 2012. Three-way complementarities: Performance pay, human resource analytics, and information technology. *Management Science* 58(5): 913–931.
- Armstrong, M. 2006. *A Handbook of Human Resource Management Practice*. London: Kogan Page Limited.
- Becker, B., Huselid, M., & Ulrich, D. 2001. *The HR Scorecard. Linking People, Strategy and Performance*. Boston: Harvard Business Press.
- Boudreau, J. 2014. Will HR's grasp match its reach? An estimable profession grown complacent and outpaced. *Organizational Dynamics* 43: 189–197.
- Chiang, R.H.L., Goes, P. & Stohr, E.A. 2012. Business intelligence and analytics education, and program development: A unique opportunity for the information systems discipline. *ACM Transactions on Management Information Systems* 3(3): Article No 12.
- CIPD 2013. *Talent Analytics and Big Data—The Challenge for HR*. London: Chartered Institute for Personnel and Development.
- Cohen, D.J. 2015. HR past, present and future: A call for consistent practices and a focus on competencies. *Human Resource Management Review* 25: 205–215.
- Corne, D., Dhaenens, C. & Jourdan, L. 2012. Synergies between operations research and data mining: The emerging use of multi-objective approaches. *European Journal of Operational Research*. 221: 469–479.
- Davenport, T.H. 2013. The Analytics Advantage. We're just getting started. Deloitte Analytics. Available at: <http://www2.deloitte.com/content/dam/Deloitte/global/Documents/Deloitte-Analytics/dttl-analytics-analytics-advantage-report-061913.pdf>. Retrieved 29.02.15.
- Demetriou, S., Kester, B., Moen B. & O'Leonard, K. 2015. HR and people analytics: Stuck in neutral. In *Global Human Capital Trends 2015*. Deloitte University Press. Available at: <http://www2.deloitte.com/content/dam/Deloitte/at/Documents/human-capital/hc-trends-2015.pdf>. Retrieved 29.02.15.
- Drucker, P.F. 1995. The information executive's truly need. *Harvard Business Review* 73(1): 54–62.
- Fitz-enz, J. 2009. *The ROI of human capital: measuring the economic value of employee performance—2nd ed*. New York: American Management Association.
- Fitz-enz, J. 2010. *The New HR analytics. Predicting the Economic Value of Your Company's Human Capital Investments*. New York: American Management Association.
- Fitz-enz, J. & Mattox, J.R. II. 2014. *Predictive Analytics for Human Resources*. Hoboken: John Wiley & Sons, Inc.
- Gantz, J. & Reinsel, D. 2012. The digital universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. Available at: <http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>. Retrieved 29.02.15.
- Gardner, N., McGranahan, D., & Wolf, W. 2011. Question for your HR chief: Are we using our 'people data' to create value? *McKinsey Quarterly*. McKinsey & Company, March: 1–5.
- Gary, M.S. & Wood, R.E. 2011. Mental models, decision rules, and performance heterogeneity. *Strategic Management Journal* 32(6): 569–594.
- IDS 2004. Searching for the magic bullet. *HR Study* 783, October: 2–6.
- Kaplan, R.S. & Norton, D.P. 2005. The Balance Scorecard. Measures that Drive Performance, Harvard Business Review. *The High Performance Organization*, July–August: 1–10.
- Kapoor, B. & Sherif, J. 2012. Human resources in an enriched environment of business intelligence. *Kybernetes* 41(10): 1625–1637.
- Kapoor, B. & Kabra, Y. 2014. Current and Future Trends in Human Resources Analytics Adoption. *Journal of Cases on Information Technology* 16(1): 1–10.
- Kark, K., White, M. & Briggs, B. 2015. 2015 CEO Global Survey. Creating legacy. Deloitte University Press. Available at: <http://www2.deloitte.com/content/dam/Deloitte/at/Documents/technology-media-telecommunications/cio-survey2015.pdf>. Retrieved 29.02.15.
- KPMG 2013. People are the real numbers. HR analytics has come of age. Available at: <https://www.kpmg.com/NL/nl/IssuesAndInsights/ArticlesPublications/Documents/PDF/Management-Consulting/People-are-the-real-numbers.pdf>. Retrieved 29.02.15.
- KPMG 2015. Evidence-based HR. The bridge between your people and delivering business strategy. Available at: <https://www.kpmg.com/Global/en/IssuesAndInsights/ArticlesPublications/Documents/evidence-based-hr.pdf>. Retrieved 29.02.15.
- Laursen, G. & Thorlund, J. 2010. *Business Analytics for Managers. Taking Business Intelligence beyond Reporting*. Hoboken: John Wiley & Sons, Inc.
- Levenson, A. 2011. Using targeted analytics to improve talent decisions. *People and Strategy* 34(2): 34–43.

- Liberatore, M. & Luo, W. 2010. The analytics movement: Implications for operations research. *Interfaces* 40(4): 313–324.
- Lim, E.P., Chen, H. & Chen, G. 2013. Business intelligence and analytics: Research directions. *ACM Transactions on Management Information Systems* 3(4), Article No 17.
- March, S.T. & Hevner, A.R. 2007. Integrated decision support systems: A data warehousing perspective. *Decision Support Systems* 43(3): 1031–1043.
- McKinsey & Company & The Conference Board 2012. The state of the human capital 2012. Research Report. Available at: <http://www.mckinsey.com/business-functions/organization/our-insights/the-state-of-human-capital-2012-report>. Retrieved 29.02.15.
- Moon, M.M. 2015a. *Talent analytics: Where are we now?* Boston: Aberdeen Group. Available at: http://v1.aberdeen.com/launch/report/research_report/10459-RR-talent-analytics-insights.asp. Retrieved 29.02.15.
- Moon, M.M. 2015b. *Five foundational metrics for meaningful workforce measurement insight*. Boston: Aberdeen Group. Available at: http://v1.aberdeen.com/launch/report/research_report/11114-RR-hr-measurement-maturity.asp. Retrieved 29.02.15.
- Moon, M.M & Prouty, K. 2015. *Productivity: Managing and measuring a workforce*. Boston: Aberdeen Group. Available at: http://v1.aberdeen.com/launch/report/research_report/10143-RR-Productivity-WFM.asp. Retrieved 29.02.15.
- Muryjas, P. & Wawer, M. 2014. Business Intelligence as a support in human resources strategies realization in contemporary organizations. *Actual Problems of Economics* 152(2): 183–190.
- Parmenter, D. 2010. *Key Performance Indicators. Developing, Implementing and Using Winning KPIs*. Hoboken: John Wiley & Sons, Inc.
- Parry, E. 2011. An examination of e-HRM as a means to increase the value of the HR function. *The International Journal of Human Resource Management* 22(5): 1146–1162.
- Pemmaraju, S. 2007. Converting HR data to business intelligence. *Employment Relations Today* 34(3): 13–16.
- Phillips, J.J. & Phillips, P.P. 2008. *Proving the value of HR: how and why to measure ROI*. Alexandria: Society for Human Resource Management.
- PwC 2015. Trends in people analytics. Available at: <https://www.pwc.com/us/en/hr-management/publications/assets/pwc-trends-in-the-workforce-2015.pdf>. Retrieved 29.02.15.
- Rasmussen, T. & Ulrich, D. 2015. Learning from practice: how HR analytics avoids being a management fad. *Organizational Dynamics* 44: 236–242.
- SAP 2010. 100 critical human capital questions—How well do you really know your organisation? Available at: <http://go.sap.com/docs/download/2015/08/2e95bcfd-377c-0010-82c7-eda71af511fa.pdf>. Retrieved 29.02.15.
- Smith, T. 2013. *HR analytics: The what, why and how*. Charlotte: Numerical Insights LLC.
- Song, L.J., Zhang, X. & Wu, J.B. 2014. A Multilevel Analysis of Middle Manager Performance: The Role of CEO and Top Manager Leadership CEO. *Management and Organization Review* 10(2): 275–297.
- Ulrich, D. & Dulebohn, J.H. 2015. Are we there yet? What's next for HR? *Human Resource Management Review* 25: 188–204.
- Watson, H.J. 2009. Tutorial: business intelligence—past, present, and future. *Communications of the Association for Information Systems* 25(1): 487–510.

The relation between friendship, commuting time, and student performance: A Social Network Analysis

L.J. Khalil & M. Khair

Notre Dame University—Louaize, ZoukMosbeh, Lebanon

ABSTRACT: Student performance is usually related both to individual-level attributes as well as to the roles of the relationships with others (such as peers, instructors or friends). Our study takes two of these factors and studies their impact on the students' GPA. Social Network Analysis (SNA) evaluates the performance of the students taking courses with a group of friends versus students used to take courses independently. A randomize model helps identifying students who choose to follow courses with friends from those who are arbitrarily in courses together. A threshold of the number of courses taken in common is used as a criterion to identify students belonging to a tribe. From the other side, the impact of the commuting distance impact on the students' GPA as well as on the drop off rate was examined. The main findings are that students in tribes over perform other students by about half point GPA, and are dropping and repeating fewer courses. Another finding is that the more a student is far in distance the less is his GPA and the more he is susceptible to drop off.

Keywords: Social Network Analysis, friendship, student performance, student retention

1 INTRODUCTION

Social networks and social network analysis has been gaining extensive importance lately and this is obvious due to the increased usage of the social network and to the opportunities for analysis and visualization that the SNA opens. SNA has been extensively used in several areas ranging from terrorism to epidemics analysis; however this is not yet extensively researched for the educational sector (Radwan 2014; Wang 2015; Della Ventura 2015). In this paper, SNA is used in order to evaluate the impact of friendship as well as the commuting time on the students' performance and the student retention.

In general, Students performance is related both to individual-level attributes (such as gender, age and socio-economic status) and to the roles of the relationships with others (such as peers, instructors or friends). Several studies have been conducted to evaluate the impact of instructor on student's performance (Kim and Sax 2011; Umbach and Wawrzynski 2005; Gasiewski et al. 2012). This study aims to determine the impact of the relations of the students on their performance by comparing their results when they took their courses individually in contrast to when they took courses with their friends. In addition, it studied the impact of the distance needed to commute to the university on the students' performance.

Based on a randomized model we separated students who are randomly allocated in courses versus those who have decided to take their courses friends. Using Social Network Analysis we created a threshold of friendship based on the number of courses taken in common.

We studied the difference in GPA of the students taking the courses alone versus those who are taking courses with friends. In addition, the impact on the drop off and the repetition of courses was evaluated.

From another perspective, we studied the impact of the commuting time on both the drop off and the students' performance.

This work identified a number of sub populations at great risk to either drop out or change major in the university. This study highlighted the importance of the social integration on the students' performance.

The first group at risk is composed by students who choose a major due to limited options ending up in a small major with few classmates (i.e. less opportunities to have a social integration), not in their field of interest.

The second group at risk is composed by students with heavy commuting time ending up in a university with few high school classmates (i.e. less contacts to have a social integration) and tired by long commuting times.

In addition, the work identified a number of means to increase the retention rates such as increasing group work to raise friendship within students in the same major.

2 LITTERATURE REVIEW

Social Network Analysis (SNA) aims to understand the determinants, the structure as well as the consequences of the relationships of individuals or groups of individuals.

2.1 *Social Network Analysis (SNA)*

The foundations of the SNA are introduced within the framework of our study, and the basic concepts as well as the terms used will be explained.

2.1.1 *SNA Node or SNA Actor*

Actors or nodes are the social units under study, which we are trying to schematize and study their associated connections. In our application, these social units are students in the educational institution. Social network perspective focuses on collections of these actors that are all of the same type, which will be referred to as one mode network and also allow focus on actors that are conceptually of different natures. For example, a two-mode network consists of two different types of actors. In our application, a two mode network consists of the students and the courses they are taking will be explained (Scott 2012, Hannemon and Riddle 2005, Prell 2012).

2.1.2 *Ties*

Ties connect actors among each other's. Ties can be unidirectional or bidirectional. Bidirectional ties means that if a tie exist between A and B, there is no difference from the meaning that the tie exists between B and A. Sometimes, however, this relation can be seen as an associated direction, such that student A likes student B, which does not imply that student B likes student A. There may be actors with dense ties and others with sparse ties which represents the relationship connections among actors (Scott, 2012, Hannemon and Riddle, 2005, Prell, 2012).

In our application, ties exist in a two-node network between students and courses, and in a one-node network, the tie exists whenever a student is taking a common course with another student.

2.1.3 *Groups*

When a tie exists between two actors, this forms a pair or dyad. In the case where the number of connected actors is more or equal than three, this is called tribe or triad. The tribe is called a clique

when each actor of a tribe is linked to all other actors of the same tribe.

2.2 *Impact of friendship on performance*

According to Carolan in (Carolan and Brian 2013), while the individual-level attributes (such as gender, age and socio-economic status) are responsible for 70 to 90% of the variation in the educational outcome, less appreciated are the roles of the relationships with others in shaping the outcomes.

Some papers try to overlook the subject and to show the number and the diversity of the research in it (Hommes et al. 2012; Biancani and McFarland 2013). Nevertheless it is important for the development of 18–22 year olds to have friends (Fullerton and Ursano 1994). This period is the end of the time that adolescents spend with their parents and spend time with their peers (Crosnoe 2000). Upon being in peer groups, students begin developing new identities, roles, and social skills (Strauss and Terenzini 2007). Those peer groups have emotional functions, as they provide social support and give a sense of belonging (Fullerton et al. 1994). Best friends provide acceptance, trust, intimacy, stability as a material, psychological, and emotional support (Cole and Bradac 1996; Richey and Richey 1980; Scholte et al. 2001).

Little work has been done to study in depth the SNA in the educational sector (Radwan 2014; Wang 2015; Della Ventura 2015) in comparison to the studies performed in other sectors. Previous authors study the influence of network association on the success, or on student's research potentials, or on student integration and persistence, or even on the distribution of knowledge (Liu and Zhu 2015; Scott 2000; Rulke and Galaskiewicz 2000; Demirbas and Demirkan 2007; Mills and Fullagar 2008).

2.3 *Impact of the commuting time on student integration and on attrition*

Commuting time can be determined as the time spent by a person to move forth and back between home and work (Choudhary et al. 2015). This determinant has been studied to evaluate its impact by (Kobus et al 2015, Mathews and Mulkeen 2005, and by Blaney and Mulkeen 2008), It was noted that the commuting distance of students to college has an impact on their integrating, and the long travel times meant they could not wait around after lectures and therefore were more likely to find it harder to make friends. In addition, it was found that university students with long commuting times were the most likely group to be non-completers when compared with other residence-types. In addition, it was noted that the more the commuting time, the less frequent the

student visits the university and the lower the students' GPA. Within the Eurostudent project (Eurostudent project), a student survey was performed in order to compare the socio-economic background and the living conditions of European students. The main output of the survey was to evaluate and compare the students commuting time and its implications.

3 AIMS OF THE STUDY

3.1 Purpose of the study

The aims of this study are to determine the impact of the relations of the students on their performance by comparing their results when they took their courses individually in contrast to when they took courses with their friends. In addition, the impact of the commuting time on the student's performance and students' retention is evaluated.

First, we defined a friendship model to separate students who are randomly in courses together and those who have decided to be with friends. SNA was used in order to evaluate and visualize the relationship between the actors. By applying the model, we defined a threshold to specify whether a student is taking courses with friends or not. This threshold is about half courses taken in common based on the distribution of common courses.

Second, we studied the difference in GPA among students in courses taken individually compared to that in courses taken in common.

Third, the impact of proximity to home on the students' success and on the students' retention is thoroughly studied by comparing the commuting time vs. the students' GPA and the relation between the drop off of students against their commuting time.

3.2 Case study

We selected a cohort of undergraduate students from six different majors in a 7000-student University. The six majors selected are chosen from different faculties and are selected out of the top fifteen enrolled majors for the years 2012–2014. The first major selected is the top one major, Bachelor of Architecture from the Faculty of Architecture. From the Faculty of Engineering, two majors are chosen, B.E. in Civil Engineering and B.E. in Mechanical Engineering. As from the Faculty of Business Administration, the Bachelor of Business Administration is chosen. From the Faculty of Natural and Applied Sciences, the B.S. in Computer Science is chosen. Lastly, B.A. in Communication Art-Radio TV is chosen from the Faculty of Humanities.

In principle, students majoring in the majors mentioned above, tend to collaborate among each

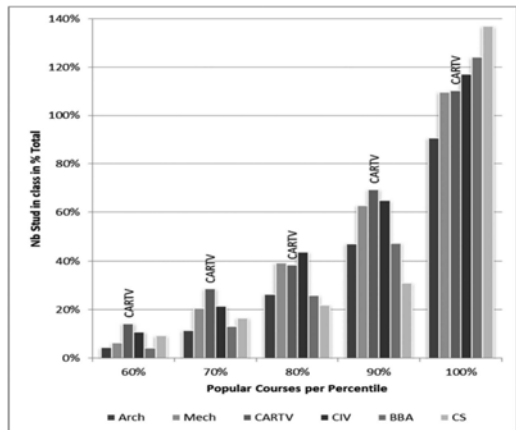


Figure 1. Popular courses per percentile.

other more than other majors (Paul and Brier 2001; Paul and Kelleher 1995). Our hypothesis is to evaluate how much the work in groups affects the whole group GPA as well as students' individual GPA.

During the 6 semesters of the study a full time student is expected to complete 24–40 courses. The number of sections of the same course given during the same semester varies from one up to 16 sections per course per semester with an average of around two sections per course. Students have bigger opportunities to define their schedule by choosing among multiple sections. Considering the respective cohort of students in the studied majors along six semesters, we found that few courses are taken in the same time by the selected cohort of students in the same major: about 10% of the courses are taken by all students and repeated, and 60% of the courses (from 0% up to 60% percentile) are taken by less than 20% of the students. (See Figure 1)

4 METHODOLOGY

To determine the elements of the tribes, the treatment of the data was done on Pajek (Nooy et al. 2011), and R. In addition, the sample size that was taken and its validity will be presented below. To calculate tribes' characteristics depending on GPA, gender, type of courses, and campus, data was processed on SPSS version 11.5 for ease and accuracy.

4.1 Construction of the network

Upon choosing students from the six selected majors as a population, we selected for each major students that have taken courses during all the six semesters inclusively from fall 2012 until spring 2015. The selected population consisted of a total

of 1,248 students distributed on the six majors as shown in Figure 7: 343 in Civil Engineering, 311 in Architecture (Arch), 259 in BBA, 199 in Mechanical Engineering, 87 in C.S., and 49 in Communication Arts-Radio/TV (CARTV) students. The next step was to manipulate the population in order to extract the friends' network.

4.1.1 From two-mode network to one-mode network

In social network analysis, matrices have been used as an efficient tool for representing a small social network. This data was presented using a binary matrix having Students as rows and Courses as columns for all the semesters, and all majors together showing the courses each student has taken (Student-Course relation).

Let M be the matrix having Students as columns and Courses as rows. The one-mode network for students is $MT * M$ and the one-mode network for Courses is $M * MT$.

4.1.2 Modeling the course selection

To build the theoretical model, we consider the weighted network of the undergraduate students of the same major in the same cohort linked by the number of courses they have taken in common during six consecutive semesters. Using the Jackknife method, we have decided to keep the real list of students and the real list of courses taken, then we allocate randomly the section of the course and the semester according to the ratio of the number of students initially in the specific session (i.e. section and semester) of the course over the total number of students enrolled in the course during the six semesters.

The created random matrix of courses students has the following properties:

- The number of students is the real number of students
- The courses taken by the students are the real courses taken by the students
- The order of the courses taken by students is random
- Session of Courses with p students will have about p students in-class.

Based on the random matrix of Courses-Students we manage through R to create the Student-students Matrix and we calculate the distribution of the number of links between students.

We repeat the random allocation of students several times (using the Monte Carlo methodology) and then we calculate the average of the distribution of the number of links between students. The Studied Network of 1,248 students has a density range from 0 up to 41 courses in common. Figure 2 shows the log-linear relation between the density and the number of vertices (students in relation). The random allocation of students gives only up to 14 courses shared

between students and the real distribution has a wide range of density from 1 up to 41 shared courses. At the level of 9 courses shared the number of links between students in the real distribution is 10 times higher than the Random distribution.

4.1.3 Threshold of friendship

The threshold of 14 courses in common has a percentage of 57% common courses. The very high density matrix with a threshold at 14 is constructed as follow: the initial matrix has been cleaned with the diagonal value down to 0 and density relations with less than the threshold have been put at 0. The very high density matrix is reduced to 243 students.

4.2 Identification of the tribes

We identified tribes and cliques or complete sub-networks based on the Pajek function Kamada-Kawai Energy. The function divides the students into Pairs, and Tribes. A network was obtained for the six majors such as the very high density network for civil engineering major shown in Figure 3.

Another method is to use R to calculate the Eigen value and vectors of the high density matrix.

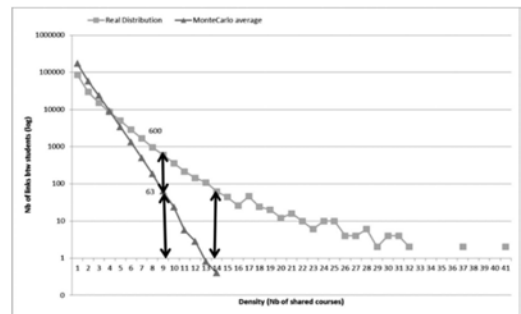


Figure 2. Log representation of density per vertices (real distribution vs Monte Carlo average).

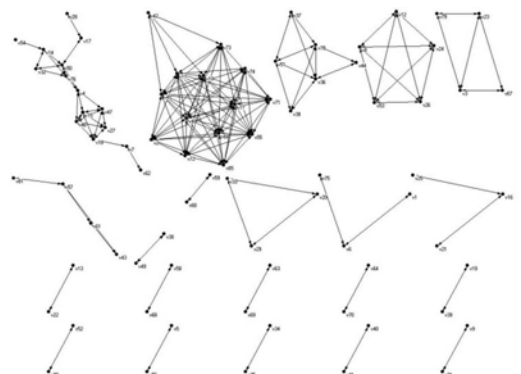


Figure 3. Tribes and pairs in civil engineering.

The Eigen vectors are the linear combination of the students in the same tribe or pair. Eigen vectors with coefficient less than 0.01 were omitted. The coefficients of the Eigen vectors are higher if the number of common courses is higher. Comparing two tribes of the same size the Eigen value is higher when the number of courses is higher.

5 FINDINGS

Tribes characteristics were analyzed depending on GPA, gender, type of courses, and campus. The following section will present the findings of the study.

5.1 Tribes characteristics

5.1.1 Campus and major impact

Half of the students in the two smaller campuses (campus 2 and campus 3) are in tribes compared to 13% in the big campus (campus 1), we can infer that students collaborate more in small campuses.

Cooperation looks more important in CARTV, Mechanical Engineering, Civil and Architecture with a ratio of about 25%. BBA and C.S. have a very low number of tribes and pairs, we think that those two majors have less cooperative project than other majors.

86% of the tribes are from different high schools especially in small campuses (goes up to 96%). This confirms the hypothesis of Shaver et al. (1985) that most of the friends in university are new friends.

5.1.2 Gender impact in tribes

The results show that females in tribes perform better in individual courses than females in pairs in all majors except C.S. and Civil Engineering. The difference is mostly significant in Mechanical Engineering where it is 0.30 for individual courses. As in common courses for females, the grade average is better in pairs than in tribes in all majors except CARTV and C.S. majors (0.15 and 0.41 respectively).

The result for males is similar in individual and common. In all the majors, except Architecture and Mechanical engineering, males are performing better in tribes. This difference is very significant in BBA and C.S. majors where the difference of average GPA is 0.94 and 0.81 for individual and common respectively in BBA and 0.80 and 0.67 for individual and common in C.S. major.

5.2 Comparing in-tribe and non-tribe students

5.2.1 Impact on Withdrawal and Repetition of courses

The percentage of withdrawal is higher in students of non-tribes, having a percentage of 15% while

tribes have 6%. The difference of percentages shows a higher difference in C.S. (16%) than other majors.

The percentage of repeating courses three times or more for all students is higher than that of students belonging to tribes.

5.2.2 Average of semester GPA in tribe and non-tribe

As an overall comparison of the average semester GPA between tribes and non-tribe students, we observe a higher GPA of students in tribes (2.86 out of 4) than students in non-tribe (2.31 out of 4). According to gender, we observe a higher average GPA in females in tribes (3.09 out of 4) than males (2.73 out of 4) in tribes, and the same for non-tribe students (2.65 in females and 2.19 in males).

Moreover, the difference in average GPA between tribes and non-tribes, according to gender, is higher in males (0.54) than in females (0.45). This shows that although females perform better in tribes, males in tribes are more efficient than females in tribes.

5.3 Impact of commuting time on students' performance

5.3.1 GPA and commuting time

There is a strong correlation between students that have low GPA (GPA 0–1) and high commuting time (35 min). As further work, we might need to analyse deeper the reasons of this correlation: one hypothesis is that they are tired commuting and it is impacting the GPA (see Fig. 4).

5.3.2 Impact on drop off students

For very low GPA (0–1) there is a 30% dropping rate, for low GPA (1–2) there is a 12% dropping rate. Drop off students with very low and low GPA have higher average commuting time (40–50 min) than average students.

For high GPA (3–4) there is 8% dropping rate. Those drop off students with high GPA have slightly the same commuting time (35 min) than average GPA students (see Table 1).

As a result we can assume only a correlation between dropping off students with very low and low GPA and commuting time.

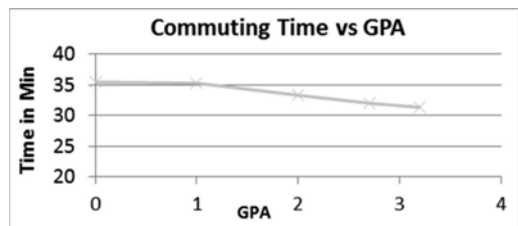


Figure 4. Relation between the GPA and the commuting time.

Table 1. Relation between the GPA, commuting time and the drop off.

GPA	Drop off rate	Average commuting time (min)
0-1	30%	46
1-2	12%	37
2-3	4%	34
3-4	8%	34

6 CONCLUSION

We define a friendship model to identify the friendship effect in a Social Network as a skew on a random distribution of the courses among students.

As a case study we consider that students are friends if they take about 50% of their courses together which means friends are students sharing more than a threshold of 14 courses out of about 30 during six semesters. This work identified a population of about 20% of the students which consider friendship as a fundamental criterion to choose a course.

The main conclusions which we can draw are the following. We confirm an already known result that most of the friends' tribes are formed by new friends who are not coming from the same high school. All students tend to have better GPA, to drop lesser and to repeat courses lesser whenever they are used to take courses with their friends.

As further work it may be of high importance to further analyze the relationships among the tribes and whether they are only in classroom or even outside by referring to the social networks and checking whether classmates' tribes continue to be friends.

REFERENCES

Biancani S., & McFarland, D. (2013). Social networks research in higher education. *Higher education: Handbook of theory and research*. Springer Netherlands, 151–215.

Blaney, C., & Mulkeen, S. (2008). Student Retention in a Modular World: A Study of Student Retention UCD Entrants 1999–2007, Retention, Modularisation, Orientation, What's Changed? Dublin: University College Dublin.

Carolan, B., & Brian, V. (2013). *Social network analysis and education: theory, methods & applications*. Sage Publications.

Choudhary, Nita, ShikhaOjha, and Niranjan Kumar Singh. (2015) "Determinants of commuting time and its impact on work and personal domains: a study measuring perception of officers in defense CPSEs." *ACADEMICIA: An International Multidisciplinary Research Journal* 5.1: 90–109.

Cole, T., & Bradac, J. (1996). A lay theory of relational satisfaction with best friends. *Journal of Social and Personal Relationships*, 13, 57–83.

Crosnoe, R. (2000). Friendships in childhood and adolescence: The life course and new directions. *Social Psychology Quarterly*, 63, 377–391.

Della Ventura, M. (2015). Music Technology: The Social Network as a Learning Resource, Modern Computer Applications. *Science and Education*, 223–228.

Demirbas, O., & Demirkan, H. (2007). Learning styles of design students and the relationship of academic performance and gender in design education. *Learning and Instruction*, 17.3, 345–359.

Eurostudent Project, http://www.eurostudent.eu/about/download_files/documents/IB_commuting_081012.pdf.

Fullerton, C., & Ursano, R. (1994). Preadolescent peer friendships: A critical contribution to adult social relatedness? *Journal of Youth and Adolescence*, 23, 43–63.

Gasiewski, J.A., Eagan, M.K., Garcia, G.A., Hurtado, S., & Chang, M.J. (2012). From gatekeeping to engagement: A multicontextual, mixed method study of student academic engagement in introductory STEM courses. *Research in Higher Education*, 53, 229–261.

Hannemon R., & Ridde M., Introduction to Social Network Methods, 2005.

Hommel, J. et al. (2012). Visualising the invisible: a network approach to reveal the informal social side of student learning. *Advances in Health Sciences Education* 17.5, 743–757.

Kim, Y.K., & Sax, L.J. (2011). Are the effects of student-faculty interaction dependent on major? An examination using multi-level modeling. *Research in Higher Education*, 52(6), 589–615.

Kobus, Martijn BW, Jos N. Van Ommeren, and Piet Rietveld. (2015). "Student commute time, university presence and academic achievement." *Regional Science and Urban Economics* 52: 129–140.

Liu, X., & Zhu, H. (2015). The Influence of Friendship Network on Graduate Student's Research Potential. In International conference on social and technology education (IC.S.STE 2015).

Mathews, N. and Mulkeen, S. (2002) 'Staying the Course? A Study of Student Retention: UCD entrants 1999–2001'. Ireland: University College Dublin.

Mills, M., & Fullagar, C. (2008). Motivation and flow: Toward an understanding of the dynamics of the relation in architecture students. *The Journal of psychology*, 142.5, 533–556.

Nooy, W., Mrvar, A., & Batagelj, V. (2011). *Exploratory Social Network Analysis with Pajek*. Cambridge University Press, New York.

Prell C, *Social Network Analysis: History, Theory and methodology*, Sage, 2013, ISBN: 978–1412947152.

Radwan, A. (2014). Evaluation of Student Performance in a Collaborative Problem Solving Environment using Social Network Media. *Recent Advances in telecommunications, informatics and educational technologies*.

Richey, M., & Richey, H. (1980). The significance of best-friend relationships in adolescence. *Psychology in the Schools*, 17, 536–540.

- Rulke, D., & Galaskiewicz, J. (2000). Distribution of knowledge, group network structure, and group performance. *Management Science*, 46.5, 612–625.
- Scott, T. (2000). Ties that bind: A social network approach to understanding student integration and persistence. *Journal of Higher Education*, 591–615.
- Scott, J. (2012). *Social network analysis: third edition*; London: Sage Publications. ISBN 978-1446209042.
- Scholte, R., Van Lieshout, C., & Van Aken, M. (2001). Perceived relational support in adolescence: Dimensions, configurations and adolescent adjustment. *Journal of Research on Adolescence*, 11, 71–94.
- Shaver, P., Furman, W., & Buhrmester, D. (1985). Transition to college: Network changes, social skills, and loneliness. In S. Duck & D. Perlman (Ed.), *Understanding personal relationships: An interdisciplinary approach* (pp. 193–219). Beverly Hills, CA: Sage.
- Strauss, L. C., & Terenzini, P. T. (2007). The effects of students' in- and out-of-class experiences on their analytical and group skills: A study of engineering education. *Research in Higher Education*, 48(8), 967–992.
- Umbach, P. D., & Wawrzynski, M. R. (2005). Faculty do matter: The role of college faculty in student learning and engagement. *Research in Higher Education*, 46(2), 153–184.
- Valente, T. (2010). *Social networks and health: Models, methods, and applications*. Oxford University Press.
- Wang, M. (2015). Social-network for Supporting Online Learning. *Computers and Technology in Modern Education*, 128–135.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Optimized clustering protocol for Wireless Sensor Networks using compressive sensing

Dina M. Omar

Department of Mathematics, University of Zagazig, Ash Sharqiyah, Egypt

Ahmed M. Khedr

Department of Computer Science, University of Sharjah, Sharjah, UAE

ABSTRACT: While Wireless Sensor Networks (WSNs) are increasingly equipped to handle more complex functions, in-network processing may require these battery powered sensors to judiciously use their constrained energy to prolong the effective network lifetime. Cluster-based Hierarchical Routing Protocol using Compressive Sensing (CS) theory (CBHRP-CS) divides the network into several clusters, each managed by a set of CHs called a headset. Each member of the head-set compresses the collected data using CS. In this paper, we propose an optimized clustering protocol using CS (OCP-CS) to improve the performance of WSNs by exploiting compressibility. In OCP-CS, each cluster is managed by a group of Cluster Heads (CHs) is called leaders-group. CHs are selected based on node concentration and sensor residual energy, and performs data aggregation using CS in order to reduce the energy consumed in the process of sampling and transmission and therefore significantly prolongs the lifetime and throughput of WSNs. Simulations show that our proposed protocol is effective in prolonging the network lifetime and supporting scalable data aggregation than existing protocols.

1 INTRODUCTION

Advances in sensor and communication technology have focused interest on using WSNs, which are formed by a set of small untethered sensor devices that are deployed in an ad hoc fashion to cooperate on sensing a physical phenomenon, making the inferences, and transmitting the data Tilak et al. (2002). WSNs can be used for a wide variety of applications dealing with monitoring (health environments, seismic, etc.), control (object detection and tracking), and surveillance (battlefield surveillance), connected cover, perimeter, and topology discovery Khedr & Osamy (2006), Khedr & Osamy (2007), Khedr & Osamy (2011), Khedr & Osamy (2012), Khedr & Osamy (2013), Khedr (2006), Khedr et al. (2009).

Energy consumption and energy-balancing are one of the primary research issues for WSNs. Since node's energy is limited and non-rechargeable, how to improve energy efficiency and balance energy has become more and more important. As the sensors in the network have limited battery power, enhancing the lifetime of a network is the basic aim of designing an energy efficient routing protocol.

Routing in WSNs is very challenging due to the essential characteristics that distinguish WSNs from other wireless networks. It is highly desirable to find the method for energy efficient route

discovery and relaying of data from sensor node to Base Station (BS) so that lifetime of network is maximized. In WSNs, the sensor nodes are often grouped into individual disjoint sets called a cluster, clustering is used in WSNs, as it provides network scalability, resource sharing and efficient use of constrained resources that gives network topology stability and energy saving attributes Kumar et al. (2011).

Clustering schemes offer reduced communication overheads, and efficient resource allocation thus decreasing the overall energy consumption and reducing the interferences among sensor nodes Khedr & Omar (2013). The basic idea of clustering routing Zhang (2009) is to use the information aggregation mechanism in the Cluster Head (CH) to reduce the amount of data transmission, thereby, reduce the energy dissipation in communication and in turn achieve the purpose of saving energy of the sensor nodes. Clustering facilitates load balancing and extends network lifetime. For example, if a CH's energy becomes depleted due to its tasks of intra-cluster communications, performing the aggregation function and inter-cluster communications, the CH may choose to resign its position; new clusters may be formed; and, other nodes may become CH to relieve the current CH of its duties. In this way, nodes in the network share the duties of being CH based on some parameter.

Accordingly, clustering strives to maximize the lifetime of the network by balancing the duties of being CH.

CS is a new sampling theory which exploits compressibility of signals in order to reduce the minimum samples required to reconstruct the original signal. Recently, CS Donoho (2006), Candes & Wakin (2008) provides a very different approach for data sampling and compression in WSNs Haupt et al. (2008), Lee et al. (2009), Bajwa et al. (2006), Jia et al. (2009) remote sensing and medical imaging. The main idea of CS is that any unknown signal X having a sparse representation in one basis (sparsifying transform) can be recovered from a small number of projections onto a second basis (sampling matrix) which is incoherent with the first one. The combination of CS theory with WSNs shows a great potential to reduce energy consumption for sensor networks which is an important factor in WSNs Candes & Wakin (2008), Zhuang et al. (2010). It reduces global scale communication cost without introducing intensive computation or complicated transmission control. This will result in extending the lifetime of the sensor network.

In this paper, we propose an Optimized Clustering Protocol using CS (OCP-CS) where, we assume that the network is randomly divided into several clusters, each managed by a group of CHs called a leaders-group and the selection of a leaders-group member is based on the residual energy and concentration degree of sensor nodes. Each member of the leaders-group compresses the collected data using CS. Simulation results show that our proposed protocol can compress data efficiently, reduce energy consumption greatly and prolonging the lifetime of the whole network to a great extent compared to other protocols.

The remainder of the paper is organized as follows: in section 2 the related work is discussed. In section 3, we present our problem statement. In section 4, we introduce the proposed system model. In section 5, we present our simulation model and analyze the 3 comparative evaluation results of the proposed protocol through simulations. And finally, conclusions are given in section 6.

2 RELATED WORK

Recently there has been a growing interest in WSNs. One of the major issues in WSN is developing an energy-efficient routing protocol. Since the sensor nodes have limited available power, energy conservation is a critical issue in WSN for nodes and network life. Heinzelman et al. (2000) proposed LEACH (Low Energy Adaptive Clustering Hierarchy) protocol, which is considered as the

basic energy efficient hierarchical routing protocol. In the setup phase of LEACH, each node decides whether to become a CH for the current round, this decision is based on a predetermined fraction of nodes and the threshold $T(s)$ as follows:

$$T(s) = \begin{cases} \frac{p_{opt}}{\left(1 - p_{opt} \times (r \bmod (1/p_{opt}))\right)} & \text{if } s \in G \\ 0 & \text{otherwise} \end{cases}, \quad (1)$$

where p_{opt} is the predetermined percentage of CHs, r is the count of current round, and G is the set of sensor nodes that have not been CHs in the last $1/p_{opt}$ rounds. Using this threshold, each node will be a CH at some round within $1/p_{opt}$ rounds. After $1/p_{opt}$ rounds, all nodes are once again eligible to become CHs. LEACH does not consider the residual energy of each node so the nodes that have relatively smaller energy remaining can be selected as CHs. This makes the network lifetime shortened. Many protocols have been derived from LEACH with some modifications and applying advance routing techniques.

Rashed et al. (2010) have proposed Cluster-Based Hierarchical Routing Protocol (CBHRP). CBHRP is an extension of LEACH Heinzelman et al. (2000). It introduces a head-set for the control and management of clusters. CBHRP divides the network into a few real clusters that are managed by a virtual CH. However, Our Proposed Protocol (OCP-CS) uses leaders-group concept and CHs are selected on the basis of their residual energy and their concentration.

CS is a sampling theory, which gives a new solution for balancing load of WSN. Based on the theory, as long as the sampled data is sparse under some basis, such as frequency domain, DFT domain, wavelet domain, it can be reconstructed through a small number of measurements with high precision Candes & Wakin (2008), Haupt et al. (2008). Using CS as the data acquisition approach in WSNs can significantly reduce the energy consumed in the process of sampling and transmission, and also lower the wireless bandwidth required for communication Zhuang et al. (2010). Under CS framework, any compressible signal $X \in \mathbb{R}^{N \times 1}$ can be represented in the form of

$$X = \Psi\alpha, \quad (2)$$

where $\Psi \in \mathbb{R}^{N \times N}$ is the transform matrix and α is the sparse representation of X . The signal X can be shown as a linear combination of K vectors with $K \ll N$, and K nonzero coefficients and $(N-K)$ zero coefficients in Equation (2). In many applications signals have only a few large coefficients. These few large coefficients can be approximated by K . One

would then select the K largest coefficients and discard $(N-K)$ smallest coefficients. Traditionally, one is required to acquire the full N -sample of signal X to compute the complete group of transform coefficients. Traditional compression techniques suffer from an important inherent inefficiency since they compute all N coefficients and records all zero coefficients, although $K \ll N$, Choi et al. (2010). The CS can replace the traditional sampling with new sampling scheme and reduce the number of measurements. In fact, CS combines the acquisition step and the compression step into one step and can directly acquire signals without going through the intermediate steps. As a result, a small number of coefficients can be transmitted or stored rather than the full set of signal coefficients. Consequently, CS provides a scheme that reduces power consumption, size and cost of the system.

The measurements of X are $Y = \Phi X$, where $\Phi \in \mathbb{R}^{M \times N}$ is a sampling matrix with far fewer rows than columns ($M \ll N$). The measurements $Y \in \mathbb{R}^{M \times 1}$ are much easier than the original networked data $X \in \mathbb{R}^N$ to be stored, transmitted, and retrieved since $M \ll N$. Therefore, the measurements can be expressed as,

$$Y = \Phi \Psi \alpha, \quad (3)$$

If $A = \Phi \Psi$ satisfies the Restricted Isometry Property (RIP) Candes & Wakin (2008) condition $M \leq cK \log(N/K)$ such that c is a small constant with $c > 0$, the vector α can be accurately recovered from Y as the unique solution of

$$\hat{\sigma} = \arg \min_{\alpha} \|\alpha\|_1 \quad s.t. \quad Y = \Phi \Psi \alpha \quad (4)$$

The original networked data X may be sparse itself or can be sparsified with a suitable transform such as Discrete Cosine Transform or Discrete Wavelet transform Lee et al. (2009). One example of the self-sparse X is a linear combination of just K basis vectors, with $K \ll N$, that is; only K s are non-zeros and $(N-K)$ s are zeros Jia et al. (2009). Usually, the networked data vector X is sparse with a proper transform in Equation (2). In WSNs, sampling matrix Φ is usually pre-designed, i.e., each sensor locally draws M elements of the random projection vectors by using its network address as the seed of a pseudorandom number generator. Based on CS theory, Jia et al. (2009) considered a sparse event detection scenario where the Channel Impulse Response (CIR) matrix is used as a natural sampling matrix. Guo et al. (2012) proposed a basic global superposition model to obtain the measurements of sensor data, where a sampling matrix is modeled as the Channel Impulse Response (CIR) matrix while the sparsifying matrix is expressed as the distributed

wavelet transform. Sartipi et al. (2011) proposed Compressive Distributed Sensing using Random Walk (CDS (RW)) for CS in WSNs that uses rateless coding. In this paper we use CS to compress data efficiently and consider residual energy and nodes concentration in CH election to achieve a robust self-configured WSN that maximizes lifetime.

Kumar et al. (2011) proposed Energy Efficient Clustering and Data Aggregation protocol for heterogeneous WSNs (EECDA). EECDA combines energy efficient cluster based routing and data aggregation for improving the performance in terms of lifetime and stability. Said et al. (2010) proposed Improved and Balanced LEACH (IB-LEACH), a heterogeneous-energy protocol examined the impact of heterogeneity of nodes, in terms of their energy, in hierarchically clustered WSNs. In these WSNs, some high-energy nodes called NCG (Normal node/Cluster Head/ Gateway) become CHs to aggregate the data of their cluster members and transmit it to the chosen "Gateways" that requires the minimum communication energy to reduce the energy consumption of CH and decrease probability of failure nodes. However, in the proposed protocol we discuss effectively the aggregation using CS and assume that CHs are randomly selected based on their residual energy and concentration degree.

As well as the differences we mentioned before, all the above protocols do not consider efficient data compression with efficient selection of CHs. In this paper we consider compression of data using CS and the selection of CHs is based on an election weight taking account of the residual energy and concentration degree of sensor nodes. Simulation shows that the proposed protocol achieves much better performance compared with CBHRP, IB-LEACH and EECDA protocols.

3 PROBLEM STATEMENT

In this paper, we propose an Optimized Clustering Protocol using CS (OCP-CS). In OCP-CS, each sensor node independently elects itself as a CH based on its residual energy and concentration degree. Each CH compresses the data received from member nodes using CS and transfers it to the BS. The proposed protocol efficiently improves data aggregation and therefore significantly reduces the energy consumed in the process of sampling and transmission and lower the wireless band width required for communication.

4 SYSTEM MODEL

In clustering, CH selection criteria strongly influence the network behavior in terms of

communication overhead, latency, and inter— and intra-cluster communication. In this paper, we propose an Optimized Clustering Protocol (OCP-CS). We assume that the CH election is based on residual energy and node concentration and use CS for aggregating data.

Our model relies on the following key assumptions regarding to the field and the sensor nodes:

1. N sensor nodes are uniformly dispersed within a square field of area $R \times Rm^2$,
2. All sensor nodes and the BS are stationary after deployment,
3. Communication is based on single-hop approach,
4. Networked data vector is sparse or highly compressible in Distributed Wavelet Transform (DWT) domain, i.e., it contains K largest coefficients. Setting the rest coefficients zero will not cause much information loss.

4.1 Architecture of OCS-CS

OCP-CS divides the network into a few real clusters. Each cluster has a leaders-group that consists of several virtual CHs; however, only one CH is active at one time. Iteration consists of two stages: an election phase and a data transfer phase. At the beginning of the election phase, a set of CHs are elected according to the residual energy and concentration of sensor nodes. These CHs send a short range advertisement broadcast message. The sensor nodes receive the advertisements and choose their CHs based on the signal strength of the advertisement messages. Each sensor node sends an acknowledgment message to its CH. Moreover, in each iteration, the CHs choose a group of associate leaders based on an election weight taking account of there sidual energy and the concentration degree of sensor nodes and the signal strength of the acknowledgments. A leaders-group consists of a CH and the associates. In the data transfer phase, the leaders-group member (CH) receives data from the neighboring nodes, aggregates the collected data using CS aiming at improving the network lifetime and reducing the network energy consumption, and then CH transmits the aggregated results to the distant BS. Finally, the BS decodes the networked data. Each data transfer phase consists of several rounds. Each member of the leaders-group becomes a CH once during a round. An epoch consistsof several iterations. In one epoch, each sensor node becomes a member of the leaders-group for one time. All the leaders-group members share the same time slot to transmit their frames. The above communication stages are illustrated in Figure 1.

4.2 Radio communication model

We use the radio energy model proposed by Heinzelman et al. (2000). According to the radio energy dissipation model illustrated in Figure 2, in order to achieve an acceptable Signal-to-Noise Ratio (SNR) in transmitting an L -bit message over a distance d , the energy expended by the radio is given by:

$$E_{Tx}(L, d) = \begin{cases} L \cdot E_{elec} + L \cdot \epsilon_{fs} \cdot d^2 & \text{if } d \leq d_0 \\ L \cdot E_{elec} + L \cdot \epsilon_{mp} \cdot d^4 & \text{if } d > d_0 \end{cases}, \quad (5)$$

where E_{elec} is the energy dissipated per bit to run the transmitter or the receiver circuit, ϵ_{fs} and ϵ_{mp} depend on the transmitter amplifier model we use, and d is the distance between the sender and the receiver. By equating the two expressions at $d = d_0$, we have $d_0 = \sqrt{\frac{\epsilon_{fs}}{\epsilon_{mp}}}$: To receive a L -bit message the radio expends $E_{Rx} = L \cdot E_{elec}$.

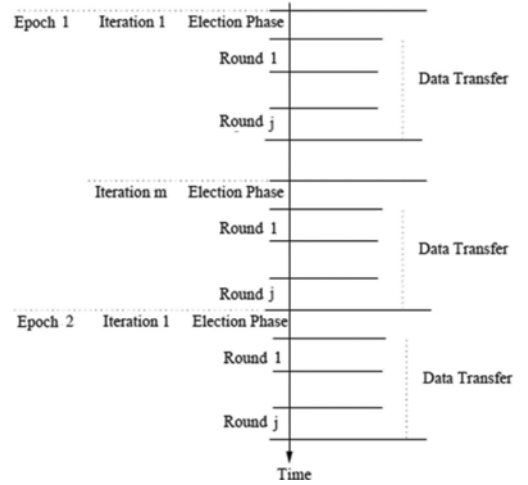


Figure 1. Communication stages in OCP-CS.

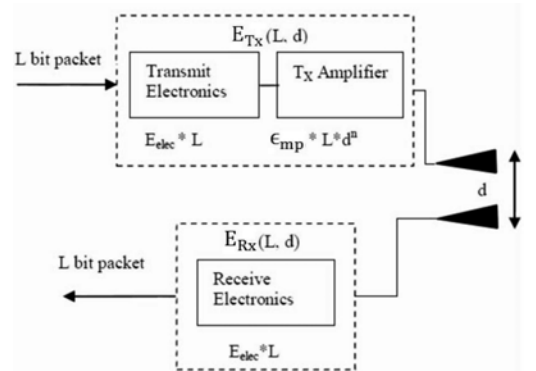


Figure 2. Radio energy dissipation model.

4.3 Optimal number of clusters

We assume that we have an area $A = R \times R$ square meters over which N nodes are uniformly distributed. For simplicity, we assume that the BS is located in the center of the field, and the distance between any node to the BS or its CH is less than or equal to d_0 . Thus, energy consumed by a CH is estimated as follows:

$$E_{CH} = \left(\frac{N}{C} - 1\right)Y.E_{elec} + \frac{N}{C}Y + Y.E_{elec} + Y\epsilon_{fs} d_{BS}^2, \quad (6)$$

where C is the number of clusters, Y is the aggregated data and d_{BS} is the average distance between CH and BS. The energy consumed by a non-CH node is given by:

$$E_{nonCH} = L.E_{elec} + L.\epsilon_{fs} d_{CH}^2, \quad (7)$$

where d_{CH} is the average distance between a cluster member and its CH. Using Euclidian metric, the area occupied by each cluster will be $A = \frac{R^2}{2\pi C}$ with node distribution $\rho(x, y)$:

$$d_{CH}^2 = \iint (x^2 + y^2)\rho(x, y)dx dy = \iint r^2 \rho(r, \theta) r dr d\theta, \quad (8)$$

Assuming the area is a circle with $\eta = R / \sqrt{\pi C}$, $\rho(r, \theta)$ is constant, and the density ρ is uniform where $\rho = (1 / (R^2 / C))$, d_{CH}^2 can be simplified as follows:

$$d_{CH}^2 = \iiint (x^2 + y^2)\rho(x, y)dx dy = \rho \int_{\theta=0}^{2\pi} \int_{r=0}^{R/\sqrt{\pi C}} r^3 dr d\theta = \frac{R^2}{2\pi C}, \quad (9)$$

The energy dissipated in a cluster per round is given by:

$$E_{cluster} \approx E_{CH} + \frac{N}{C}E_{nonCH}, \quad (10)$$

The total energy dissipation in the network per round will be the sum of the energy dissipation by all clusters, i.e.,

$$E_{tot} = CE_{cluster} = Y\left(N(1 + E_{elec}) + C\epsilon_{fs} d_{BS}^2\right) + NL(E_{elec} + \epsilon_{fs} d_{CH}^2), \quad (11)$$

By differentiating E_{tot} with respect to C and equating to zero, the optimal number of constructed clusters can be found:

$$C_{opt} = \sqrt{\frac{NL}{2\pi Y}} \frac{R}{d_{BS}} = \sqrt{\frac{NL}{2\pi Y}} \frac{2}{0.765}, \quad (12)$$

where, the average distance from a CH to the BS d_{BS} is given by Bandyopadhyay et al. (2004) as follows:

$$d_{BS} = \int \sqrt{x^2 + y^2} \frac{1}{A} dA = 0.765 \frac{R}{2}, \quad (13)$$

If the distance of a significant percentage of nodes to the BS is greater than d_0 then, following the same analysis as in Heinzelman et al. (2002) we will obtain:

$$C_{opt} = \sqrt{\frac{NL}{2\pi Y}} \sqrt{\frac{\epsilon_{fs}}{\epsilon_{mp}}} \frac{R}{d_{BS}^2}, \quad (14)$$

The optimal probability of a node to become a CH, p_{opt} , can be computed as follows:

$$p_{opt} = \frac{C_{opt}}{N}, \quad (15)$$

The optimal construction of clusters is very important. Heinzelman et al. (2000) showed that if the clusters are not constructed in an optimal way, the total consumed energy of the sensor network per round is increased exponentially either when the number of the constructed clusters is greater than the optimal number of clusters or especially when the number of the constructed clusters is less than the optimal number of clusters. If the number of the constructed clusters is less than the optimal number of clusters, some nodes in the network have to transmit their data very far to reach the CH, causing the global energy in the system to be large. If the number of the constructed clusters is greater than the optimal number of clusters, the total routing traffics within each cluster will be reduced because of fewer members, however, more clusters will result in more one-hop transmissions from the CHs to the BS also the CHs will receive data from fewer members this will reduce the local data aggregation being performed and increase the communications among the CHs.

4.4 Cluster head election phase

The optimal probability of a node to become a CH is equivalent to the optimal construction of clusters. This clustering is optimal in the sense that energy consumption is well distributed over all sensors and the total energy consumption is minimal. Such optimal clustering highly depends on the energy model that we use.

Energy consumption of the CHs is relatively expensive, so the residual energy of sensor node is the main criteria for the election of CH. Moreover, data aggregation can save considerable energy when the source nodes forming one cluster are distributed in a relatively small region while the BS is far away from the source nodes, because sensor nodes only need much few energy for sending data to the CH than sending data directly to the BS when the BS is located at a remote distance Krishnamachari et al. (2002). So it is reasonable to infer that the closer source nodes within a cluster, the lower energy they need to consume to send data.

On account of the deduction above, an election weight taking account of the residual energy and the concentration degree of sensor nodes is introduced in OCP-CS for CH election.

Definition 1: Given a WSN of N sensor nodes ($i = 1, 2, \dots, N$), $D^r(i)$ is defined to be the concentration degree of node i , namely the number of sensor nodes that can sense during the r^{th} round. $W(i, r)$ is defined as the election weight of node i in r^{th} round,

$$W(i, r) = \omega \frac{E_i^r}{\bar{E}^r} + (1 - \omega) \frac{D^r(i)}{p_{opt}}, \quad (16)$$

where $\omega = \frac{1}{1+\xi}$ is an adaptive factor to adjust the impact of residual energy and concentration degree on the election weight, $\xi = \frac{E_i^r}{\bar{E}^r}$ denotes the residual energy of node i in round r , E_i^r is the initial energy of node i and \bar{E}^r is the average residual energy of network in r^{th} round. With the reduction of residual energy, ω will gradually increase to adapt to the decrease of the number of effective sensor nodes in WSN.

4.5 Setup phase

- *Step 1* During the initialization, sensor nodes calculate their own concentration degree according to Definition 1, and mark their own level as level 1.
- *Step 2* In the initialization phase of the network, the BS broadcasts \bar{E}^r in “CH election” messages. When a node i receives the broadcast message, it will compare its own residual energy E_i^r with \bar{E}^r . If $E_i^r \geq \bar{E}^r$, a node i will calculate the election weight using its own $D^r(i)$ and E_i^r , and then send the weight and its ID to the BS for CH election in the “CH election” messages. Otherwise, a node i will give up CH election, and chooses to join a cluster later.
- *Step 3* The BS marks its own level as level 1, chooses C_{opt} sensor nodes with maximum election weight as CHs. Sensor nodes that have been chosen to be the CHs by the BS will mark

themselves as CHs. After that, CHs will broadcast to neighbor nodes to notify them that it has been elected as a new CH.

- *Step 4* When a node is elected as a CH, it will broadcast “re-join the cluster” messages to each regular node. After receiving the broadcast message, each regular node chooses its closest CH with the largest received signal strength and then informs the CH by sending a join cluster message. Furthermore, in each iteration, the CHs choose a set of associate leaders based on their election weight and the signal strength of their acknowledgments. A leaders-group consists of a CH and the associates. The leaders-group member is responsible for sending messages to the BS.
- *Step 5* The CH sets up a Time-Division Multiple Access (TDMA) schedule and transmits it to the nodes in the cluster. After the TDMA schedule is known by all nodes in the cluster, the set up phase is completed and the next phase begins.

4.6 Data transmission phase

Once the clusters are formed and the TDMA schedule is fixed, the data transmission phase can begin. We consider N sensors randomly located in a field, each generating a data sample x_j ($j = 1, \dots, N$) to be measured. The vector of data samples $X = [x_1, \dots, x_N]$ is called networked data Haupt et al. (2008), which will be transmitted to the BS. We use Discrete Wavelet Transform (DWT) matrix as the sparsifying transform matrix and Channel Impulse Response (CIR) matrix as the sampling matrix.

4.6.1 DWT basis

We assume that the sensed data is highly correlated in space domain. We use Discrete Wavelet Transform (DWT) to sparsify the networked data X and DWT is applied to the sampled data. DWT attempts to de-correlate the correlated data into uncorrelated coefficients using a group of wavelet basis functions. Once the BS knows the locations of all sensor nodes, DWT basis can be computed. DWT replaces the 2-D set of measurements with a set of transform coefficients that, for piecewise smooth fields, are sparser than the original data,

$$X = WS, \quad (17)$$

where $S \in \mathbb{R}^N$ is the transform coefficient vector which contains K ($K \ll N$) nonzeros, and W is the DWT basis.

4.6.2 CIR basis

At each cluster, a leaders-group member CH receives data from the neighboring nodes, aggregates the

collected data using CS and transmits the aggregated results to the distant BS. The received signal vector at CH can be written as,

$$Y = GX = GWS, \quad (18)$$

where G is the CIR matrix whose component can be written as,

$$G[m, n] = d_{m,n}^{-\gamma} |h_{m,n}|. \quad (19)$$

where $d_{m,n}$ is the distance between the m^{th} CH and the n^{th} sensor node. γ is the propagation loss factor

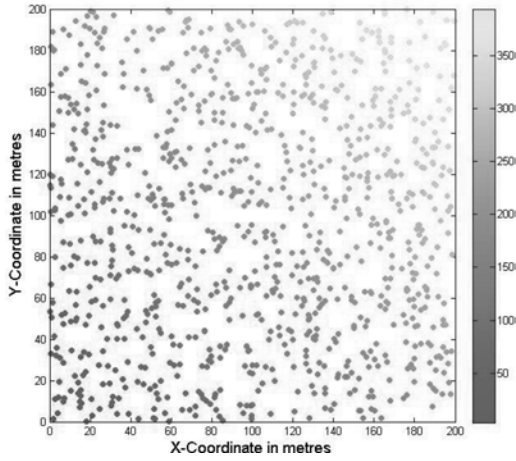


Figure 3. Network topology comprised of 1000 nodes. Note: The locations of nodes are generated as random values drawn from the standard uniform distribution on the open interval (0, 200).

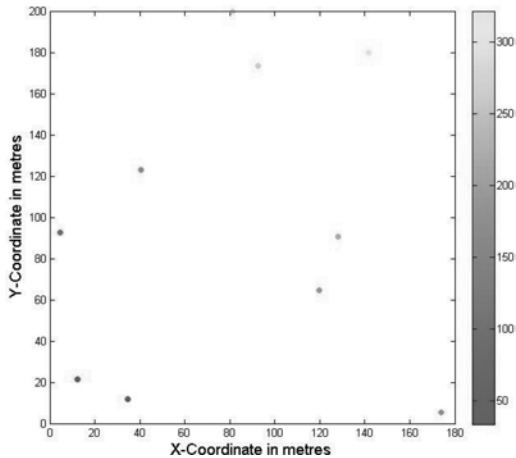
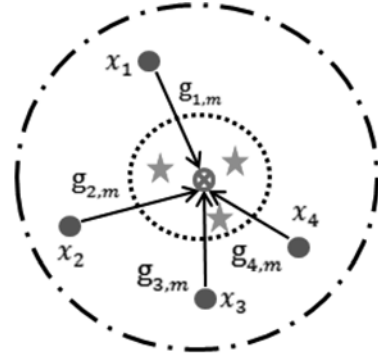


Figure 4. Sparsity of networked data in a DWT basis. The networked data can be presented with $K = 10$ nonzero coefficients after DWT transform.



- ⊗ Cluster Head
- ★ Member of Head-Set
- Member Node

Figure 5. Transmission in clusters.

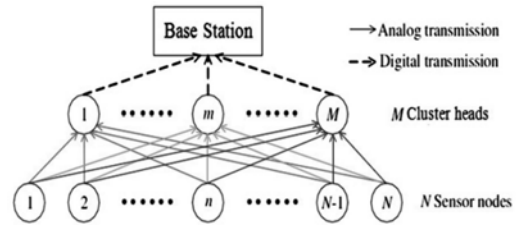


Figure 6. The basic CIR model.

which is 2 for free space Gay-Fernandez et al. (2010) and takes on other values for different media Jia et al. (2009). $h_{m,n}$ is the Rayleigh fading coefficient modeled as complex Gaussian noise with zero mean and unit variance Jia et al. (2009).

As shown in Figure 6 N sensor nodes transmit their samples to M CHs. Subsequently CHs transmit measurements Y to the BS independently. Finally, the BS decodes the networked data X from Y using for example the basis pursuit solver in Sparselab toolbox of MatlabDonoho (2006).

5 SIMULATION RESULTS

In this section, the analysis of the proposed OCP-CS protocol is carried out using MATLAB to evaluate the energy consumption and maximize the lifetime of the sensor network. We describe the simulation environment, performance metrics and experimental results.

The simulation parameters are summarized in the following table (Table 1):

Table 1. Simulation parameters.

Description	Parameter	Value
No. of nodes	N	1000
Initial energy	E_0	0.5
Location of the BS	BS	(50,50)
Data packet size	L	4000 bits
Network area	$R \times R$	$200 \times 200 m^2$
Transmit amplifier	ϵ_{fs}	$10 pJ/(bit \cdot m^2)$
If $d_{BS} \leq d_0$		
Transmit amplifier	ϵ_{mp}	$0.0013 pJ/(bit \cdot m^4)$
If $d_{BS} \geq d_0$		
Threshold distance	d_0	87.7058 m
No. of nonzero	K	10
Coefficients		
No. of measurements	M	50
Propagation loss factor	γ	2

5.1 Performance metrics and experimental results

Here, we present the performance results and the comparison of the proposed protocol with other existing protocols. The evaluation of performance metrics demonstrate the improvement and strength features of the proposed protocol compared with EECDA, IB-LEACH and CBHRP protocols.

5.1.1 Energy consumption

Since energy consumption is the most important issue in WSNs, we discuss the impact of using CS on energy consumption by comparing the performance of the proposed protocol with existing protocols. The energy consumption for specific number of frames with respect to the variation of cluster number and network diameter size is examined.

Figure 7 illustrates the difference of the energy consumed per round in the proposed, EECDA, IB-LEACH and CBHRP protocols. It shows that IB-LEACH achieves better performance compared with CBHRP, whereas the gateways take up the role to reduce the energy consumption of CH and decrease probability of failure nodes. Also, EECDA performs better than IB-LEACH, the reason is EECDA selects path with a maximum sum of energy residual for data transmission in spite of that path with minimum energy. It is obvious that the energy consumption of the proposed protocol is much lower than that of CBHRP, IB-LEACH and EECDA. This is because OCP-CS uses an election weight taking account of the residual energy and the concentration degree of sensor nodes in electing CHs; node having higher election weight has greater chances to be a CH, therefore, the energy efficiency is enhanced. Besides, OCP-CS efficiently compresses data and at the same time guarantees fast data compression which is an important issue in WSNs due to the scarce resources of sensor node.

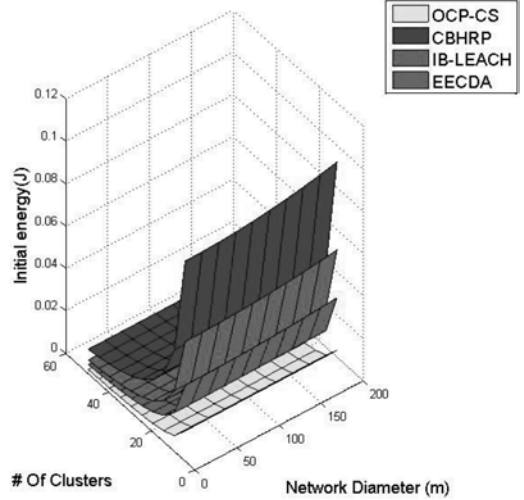


Figure 7. Energy consumption.

Consequent to this compression, the total network energy consumption is minimized compared with CBHRP, IB-LEACH and EECDA. Figure 7 shows that energy consumption is reduced when the number of clusters increases and the network diameter decreases. For the simulated network of 1000 nodes, it is shown that the optimum range of clusters lies between 20 and 60. As the number of clusters increases and the network diameter decreases, the energy consumption also decreases. When the number of clusters is below the optimum range, for example 10, the sensor nodes have to send data to the distant CHs. On the other hand, when the number of clusters is greater than optimum range, there will be more transmissions to the distant BS. Moreover, when the network diameter increases, the CHs have to send data to the distant BS. Furthermore, when the network diameter decreases, the energy consumption also decreases and there will be more transmissions to the BS.

5.1.2 Iteration time

In WSNs, the most important metric is the total survival lifetime of the network. In this section, the average time to complete one iteration such that every node becomes a member of leaders-group is analyzed using OCP-CS and compared with existing protocols.

The estimated time for one iteration with respect to the network diameter considering the percentage of leaders-group size is shown in Figure 8. It is obvious that the estimated time for one iteration of the proposed protocol is more than that of EECDA, IB-LEACH and CBHRP. Whereas, in the proposed protocol the extension of the network service duration is because OCP-CS efficiently compresses data

using CS and every sensor node independently elects itself as a CH based on its election weight. Therefore, OCP-CS would extend the estimated time for one iteration, and consequently the battery lifetime would be extending to more than current lifetime. The iteration time is proportional to the initial energy and the network diameter found in this figure. The network will be alive for a longest period of time with initial energy when the leaders-group size is 50% of the cluster size. However, it is more or less with respect to the leaders-group size.

Figure 9 shows a graph that illustrates the estimated time for one iteration with respect to the number of clusters and leaders-group size. The figure shows that for the same number of clusters, the time for iteration increases as the leaders-group size increases and one iteration can last longer for larger leaders-group sizes. However, for larger number of clusters, the time for iteration is reduced. This graph shows that the leaders-group size and the number of clusters should be carefully chosen to extend the network lifetime. The figure shows that OCP-CS outperforms EECDA, IB-LEACH and CBHRP protocols, the reason are using CS would optimize energy usage and this leads to prolonging the network lifetime.

5.1.3 Number of frames

The number of frames transmitted in each iteration is evaluated using OCP-CS and compared with EECDA, IB-LEACH and CBHRP.

Figure 10 shows the number of frames transmitted per iteration in the proposed, EECDA, IB-LEACH and CBHRP protocols. It is clear that the proposed protocol outperforms existing protocols. Also, it is shown that when the leaders-group size increases, there are more control and management sensor nodes. As a result, the iteration can last for a longer time, which is also consistent with the results

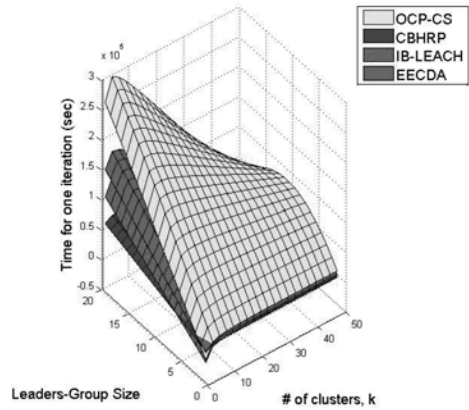


Figure 9. Time for iteration with respect to number of clusters and leaders-group size.

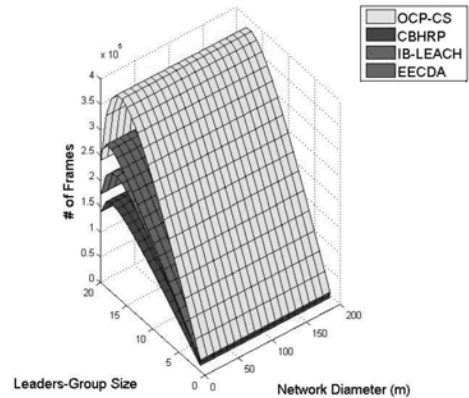


Figure 10. Number of frames transmitted per iteration.

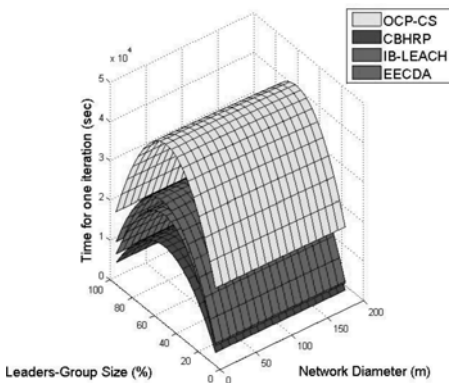


Figure 8. Time for iteration with respect to network diameter and leaders-group size.

shown in Figure 8 and Figure 9. Consequently, the data collecting nodes can be used for a longer period of time. Our results show that the proposed protocol provides a more systematic approach of transmitting a higher number of data frames in contrast to EECDA, IB-LEACH and CBHRP.

6 CONCLUSION

In this paper, we propose an optimized clustering protocol using compressive sensing to enhance the energy consumption, lifetime and throughput of the wireless networks. Compressive sensing measurements are obtained via cluster heads. Discrete Wavelet Transform (DWT) is used as the sparsifying matrix and Channel Impulse Response (CIR) matrix is used as the sampling matrix. Using leaders-group concepts in a clustering algorithmic

approach, nodes elect themselves as cluster heads based on their energy levels and concentration degree, retaining more uniformly distributed energy among sensor nodes. The simulation results show that our protocol decreases the energy consumption and therefore, prolongs the network lifetime and increases the number of frames transmitted per iteration compared with EECDA, IB-LEACH and CBHRP protocols.

REFERENCES

- Bajwa, W., Haupt, J., Sayeed, A., and Nowak, R. (2006). Compressive wireless sensing. *5th Int. Conf. Information Processing in Sensor Networks*. 134–142.
- Bandyopadhyay, S., Coyle, E.J. (2004). Minimizing communication costs in hierarchically-clustered networks of wireless sensors. *Computer Networks*. 44(1), 1–16.
- Candes, E.J. and Wakin, M.B. (2008). An introduction to compressive sampling. *IEEE Signal Process. Mag.* 25(2), 21–30.
- Choi, K., Wang, J., Zhu, L., Suh, T.S., Boyd, S. and Xing, L. (2010). Compressed sensing based cone-beam computed tomography reconstruction with a first-order method. *Med. Phys.* 37, 5113–5125.
- Donoho, D.L. (2006). Compressed sensing. *IEEE Trans. Inf. Theory*. 52(4), 1289–1306.
- Gay-Fernandez, J.A., Sanchez, M.G., Cuinas, I., Alejos, A.V., Sanchez, J.G., and Miranda-Sierr, J.L. (2010). Propagation analysis and deployment of a wireless sensor network in a forest. *Progress in electromagnetics research*. 106, 121–145.
- Guo, D., Qu, X., Huang, L., Yao, Y. (2012). Optimized local superposition in wireless sensor networks with T-average-mutual-coherence. *Progress in electromagnetics research*, 122, 389–411.
- Haupt, J., Bajwa, W.U., Rabbat, M. and Nowak, R. (2008). Compressed sensing for networked data. *IEEE signal process Mag.* 25(2), 92–101.
- Heinzelman, W., Chandrakasan, A. and Balakrishnan, H. (2000). Energy-efficient communication protocol for wireless microsensor networks. *Proceedings of 3rd Hawaii international conference on system sciences*.
- Heinzelman, W.B., Chandrakasan, A.P., Balakrishnan, H. (2002). An Application-Specific Protocol Architecture for Wireless Microsensor Networks. *IEEE Transactions on Wireless Communications*. 1(4), 660–669.
- Jia, M., Husheng, L. and Zhu, H. (2009). Sparse event detection in wireless sensor networks using compressive sensing. *3rd Annu. Conf. information sciences and systems*. 181–185.
- Khedr, A.M. (2006). Tracking Mobile Targets using Grid sensor networks, *GESJ: Computer science and Telecommunications*, 3(10), 66–84.
- Khedr, A.M. and Omar D.M. (2013). SEP-CS: Effective Routing Protocol for Heterogeneous Wireless Sensor Networks, *Ad Hoc & Sensor Wireless Networks*, 0, 1–22.
- Khedr, A.M. and Osamy W. (2006). A Topology Discovery Algorithm For Sensor Network Using Smart Antennas, *Computer Communications Journal*, 29, 2261–2268.
- Khedr, A.M. and Osamy W. (2007). Target tracking-Mechanism for Cluster Based Sensor Networks, *Applied Mathematics and Information Science Journal*, 1(3), 287–303.
- Khedr, A.M. and Osamy W. (2011). Effective Target Tracking Mechanism in a Self-Organizing Wireless Sensor Network. *Journal of Parallel and Distributed Computing*, 71, 1318–1326.
- Khedr, A.M. and Osamy W. (2012). Mobility-assisted minimum connected cover in a wireless sensor network, *J. Parallel Distrib. Comput.* 72, 827–837.
- Khedr, A.M. and Osamy W. (2013). Minimum Connected Cover of Query Regions in Heterogeneous Wireless Sensor Networks, *Information Sciences*, 223, 153–163.
- Khedr, A.M., Osamy W. and Agrawal D.P. (2009). Perimeter Discovery in Wireless Sensor Networks, *J. Parallel Distrib. Comput.* 69, 922–929.
- Krishnamachari, B., Estrin, D., Wicker, S. (2002). The Impact of Data Aggregation in Wireless Sensor Networks. *Proceedings of International Workshop on Distributed Event-Based Systems*.
- Kumar, D., Aseri, T.C., Patel, R.B. (2011). EECDA: Energy Efficient Clustering and Data Aggregation Protocol for Heterogeneous Wireless Sensor Networks. *Int. J. of Computers, Communications & Control*. 6(1), 113–124.
- Kumar, S.V., Jain, S. and Tiwari, S. (2011). Energy Efficient Clustering Algorithms in Wireless Sensor Networks: A Survey. *IJCSI International Journal of Computer Science Issues*. 8(2), 1694–0814.
- Lee, S., Patten, S., Sathiamoorthy, M., Krishnamachari, B., and Ortega, A. (2009). Spatially-localized compressed sensing and routing in multi-hop sensor networks. *3rd Int. conf. on Geo. Sensor networks*. 11–20.
- Rashed, M.G., Kabir, M.H., Rahim, M.S. and Ullah, S.E. (2010). Cluster Based hierarchical routing protocol for wireless sensor network. *International Journal of computer and network security*. 5(2).
- Said, B-A., Abdellah, E., Beni-Hssane, A., and Hasnaoui, M. L. (2010). Improved and Balanced LEACH for heterogeneous wireless sensor networks. *International Journal on Computer Science and Engineering*. 2(8), 2633–2640.
- Sartipi, M. and Fletcher, R. (2011). Energy-efficient data acquisition in wireless sensor networks using compressed sensing. *Data compression conference (DCC)*. 223–232.
- Tilak, S., Abu-Ghazaleh, N., and Heinzelman, W. (2002). A Taxonomy of Wireless Microsensor Network Models. *Computer Journal of ACM Mobile Computing and Communication Review (MC2R)*. 6(8), 1–8.
- Zhang, Z. and Zhang, X. (2009). Research of Improved Clustering Routing Algorithm Based on Load Balance in Wireless Sensor Networks. *IET International Communication Conference on Wireless Mobile and Computing*. 661–664.
- Zhuang, X., Houjun, W., and Zhijian, D. (2010). Wireless sensor networks based on compressed sensing. In *3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT2010)*. 9, 90–92.

Compositional writing to cope with electronic media

Fodil Mohammed Sadek

Department of English, Faculty of Letters and Languages, Mouloud Mammeri University of Tizi-Ouzou, Tizi Ouzou, Algeria

ABSTRACT: This paper examines the relationship between the explosion of knowledge today, and the lack of appropriate cognitive means to sustain it. Historically, and from the linguistic standpoint, an outstanding progress was made when men invented writing as a means to respond to the manifold communication needs of the Sumerians who awfully sought a safe device to store their memories and knowledge which reached such amounts that no human memory could preserve. By projecting thought and knowledge outside the human body, writing set the ground for literacy. Knowledge progressed and writing served as the dominant medium to both preserve and share this knowledge over long distances. The development of printing four thousand years later considerably improved distant communication. The book became the standard means for communication interchange, and, in so doing, printing also imposed a novel way of conceiving writing by promoting editorial constraints which have eventually impacted the way people organize their thoughts before engraving them side by side, one after another, in a limited physical space mainly consisting in a standard white page. The advent of the internet has considerably increased human communication potential by offering a non-linear writing space. A text editor may also resemble a standard white page, but it is formatted in such a way as to integrate aside ordinary alphabetic writing, the use of hypertext and hypermedia, two of the most common adopted electronic writing means which have augmented human intelligence. However, despite the availability of these electronic tools which have the potential to initiate a real shift in knowledge representation, the traditional type of linear writing using alphabetical systems remains dominant, mainly in scholarly production. What is observed in the field is the absence of appropriate non linear linguistic structures liable to answer the new communicative need. Therefore, the linguistic challenge today is to extend Humboldt's definition of language as a system which makes 'infinite use of finite means', by integrating novel symbols in the construction of neologisms able to convey different levels of meanings. In this respect, it will be argued that componyms appear as relevant structures likely to assume this role.

Keywords: knowledge; communication; writing; internet; hypertext; componyms

1 INTRODUCTION

Any human gathering requires an efficient communication tool to maintain safe the cohesion of the group. Scholars consider that human linguistic communication began with speech, despite a mimetic communication that preceded it. The latter view defended by scholars like Logan¹ considers that speech emerged opportunely to help hominids conceptualize thought. This important step was made possible by the move from perception to conception as a way to cater for an information overload resulting from the developing minds of hominids. To Logan², language represents a bifurcation from mimetic to verbal communication

when the amount of information to communicate exceeded by far the possibilities made available by mimetic communication. In other words, language arose as a transition from percept-based thinking to concept-based thinking, and this move promoted their evolution from Hominids to Homo sapiens.

In this frame of thought, speech emerged as an appropriate answer to the new communication needs experienced by Homo sapiens, whose social life became more complex, thus requesting more elaborated communication means. Speech helped regulate verbal interaction by ensuring a face to face type of communication based on a common code. The code, though basic, necessarily involved rules (phonological, grammatical, semantic and pragmatic), and it was sufficient for humans to serve as a reliable system for the representation of abstract knowledge, and to cater for their communication needs until the invention of writing and the development of literacy.

¹Logan, R.K. (2006). The extended mind model of the origin of language and culture. In: N. Gontier, J.P. Van Bendegem and D. Aerts (Eds.) *Evolutionary epistemology, language and culture*. Dordrecht. Springer.

²Idem.

By deciding to use written graphs to represent elements of thought, the first Sumerians invented a semiotic practice that considerably formatted the later expression of meaning. They made use of means external to the human body, like clay tablets made from a combination of a piece of earth and water moulded into geometric shapes so to serve as a locus ready to receive specific markings impacted by a stylus. Consequently, writing evolved into a new system of representation of abstract knowledge. As such, it became a practical medium to help devise methods to conceive and organize information by projecting thought outside the human body into some sort of physical space where information could be organised in explicit ways, be transported to distant areas, and be endlessly disclosed on demand. In short, writing based on a linear combination of graphs imperfectly representing sounds, allowed for an asynchronous type of communication over long distances. Paradoxically, as will be shown further, this linear space has also become its prison in the long run.

It is commonly agreed that the divide between humans and non-humans started with the use of natural language based on the double articulation of language. Human language functions as an interface for the representation of both the outer world of things, and the inner world of thought thanks to the double articulation of language. This consists, according to André Martinet³, for a speaker to first articulate their inner thoughts into meaningful units called monemes. The monemes are in their turn articulated into smaller units called phonemes. The phonemes coming out of the mouth in the form of oral sounds are addressed to a listener who shares the same code, and who, under ideal circumstances understands the streams of sounds as corresponding to particular units of meaning. In this case the message intended by the speaker achieves its purpose and communication succeeds. In the case of writing, the oral sounds are transcribed into written characters corresponding roughly to the sounds of the language, according to specific spelling rules.

This double articulation has always catered for all the communicative needs of humans using alphabetic types of languages, until the appearance of computer-mediated communication and particularly of networked computers. In effect, the explosion of knowledge following the arrival of cyberspace has brought into relief the limits to the creativity of human language, due to the linearity of language which originates precisely in the double articulation. Huge amounts of information ought to be conveyed simultaneously to several

distant places but using linguistic tools that are not suitable for such functions. This point will be fully developed after setting a historical perspective on the limits of human communication.

Innis and Mc Luhan⁴ divided the history of human communication into 3 main ages, namely the age of orality, the age of literacy, and the age of electric mass media, while Danesi⁵ omits the first age to focus only on the other two. Indeed, one of the assets of the printed word was to let an author construct meaning, starting from a white page which a series of previously arranged graphs progressively darken to confer the page a particular graphic architecture. The wording of the writing/graphic space is linear, and it is exploited by an author in such a way as the graphic arrangements result for the reader in an understanding as close as possible to the one expected by the author. Accordingly, it is the salient properties of the graphic architecture specifically marshalled by the author which orient the reader's reflection towards a kind of understanding rather than towards another, less expected by the writer. The words are there on a page, and their particular positions and agencies within the white page produce specific effects on the minds of the readers. As Eisenstein⁶ clearly points out, 'The development of print was significant in that it reinforced the linearity and sequentiality of writing while focusing on the hierarchical thinking that was essential to the eventual flowering of modern science'.

This wording pattern became a stable layout for the standard printed text whatever the specificities of the language concerned. Nevertheless, human thinking and cognitive processes resulting from neuronal activity are far from being linear. As a matter of fact, and despite the ingeniousness of the double articulation of language on the one hand, and of the editorial devices imposed by the printing industry to cater for the complexity of human thought now projected into a flat paper space, time came when these tools proved unfit to account for such a complexity induced by today's communication needs, because human thought remains basically dynamic and prefers networks to single linear paths.

To the three above mentioned ages, Logan⁷ added two others. The first he called the age of mimetic communication (archaic *Homo sapiens*)

⁴Mc Luhan, M. *Understanding Media, The Extensions of Man*, The MIT Press, 1994.

⁵Marcel Danesi, "Understanding Media Semiotics" OUP, 2002.

⁶E. Eisenstein. *The printing revolution in early modern Europe*. Cambridge, UK: Cambridge University Press. 1983.

⁷Logan, Robert. *The alphabet effect: The impact of the phonetic alphabet on the development of Western civilization*. New York: William Morrow (1986).

³Martinet, André. 1998. *Éléments de linguistique générale*. Armand Colin.

which preceded speech, and the last he termed the age of digital interactive media. The focus will be put on this interactive aspect of digital media which is the major development brought by ICTs to the communication era. In effect, never in history was man able to connect with distant others so instantly, merging both synchronous and asynchronous types of multimedia communication, thus offering readers several possibilities to seek or track information via hypertexts and hypermedia, freeing them from the constraining linearity of print writing. In other words, the new literacy age has started with the development of network thinking, and as Danesi⁸ points out: “any major change in how information is represented and transmitted brings about a concomitant paradigm shift in cultural systems.” It makes no doubt that networked thinking calls for subsequent network structures to construct novel meaning Ferris⁹ has already mentioned that ‘Computers incorporate a new orality by bringing new perspectives to the manipulation and understanding of writing. The text becomes more immediate, more fragmented and fluid, and the medium offers greater capacity for individual participation and interactivity’.

The interactivity of the new information and communication technologies was made possible by human communication mediated by the computer. An interesting offshoot of this interactivity is the multi-layered type of writing consisting in an original manner of articulating meaning assumed both by hypertext and at a lesser extent, by componyms¹⁰. Componyms which will be illustrated below, appear as suitable linguistic structures liable to be integrated in virtual spaces to ordinary linear writing, because they offer thought unlimited possibilities to construct novel meaning beyond the limits of alphabetic writing. This inventive writing is fuelled by the adoption of innovative lexicogenic devices involving graphs, digraphs, and alphanumeric characters in addition to other signs like @. Indeed, network thinking, untied from the linearity of the physical constraints of the classical type of writing from left to right or right to left, on the one hand, and from the tiny contiguity of alphabetic writing imposed by phonological systems, is now facing a wide area of possible paths likely to embark human communication on new intellectual challenges.

⁸Marcel Danesi, “Understanding Media Semiotics” OUP, 2002.

⁹Shamila Pixy Ferris. Writing Electronically: The Effects of Computers on Traditional Writing.

¹⁰See complete article in Fodil, M.S. MICUs, Componyms and the Triple Articulation of Cyber Language. <http://revue.ummto.dz/index.php/khitab/article/view/814/654>.

The new task set for linguists today, is to contribute to transform the present information chaos to ordered structures. The information chaos results from a rapid shift within a few decades only, from a situation where humans lacked information, to another where the over abundant sources of information coupled with unprecedented facilities to disseminate it, produce what some call “info-pollution”. Indeed, scholars are urged to conceive the theoretical and methodological tools that will help order the overwhelming amounts of information and knowledge available in a more or less chaotic shape, into legible linguistic structures liable to endow them with an order that fits the demands of the Twenty First Century.

One of the possible solutions is to devise novel linguistic structures able to convey more complex meaning by making use of novel lexicogenic processes known as componyms. Indeed, componyms resemble ordinary simple lexical units such as nouns, verbs, etc., but while simple lexical items are composed of phonemes, componyms are built from initials of words which, as a solitary whole, compose a complex structure. For example, *to grep*, stands for: (*globally search a regular expression and print the lines containing matches to it*). The novelty with this type of lexicogenic processes is to augment the loading capacity of words by adding other layers of meaning to their representational linguistic labels. The effect produced by componyms like *to grep*, is that what looks at first as an ordinary word turns out to be a whole sentence expressed in the imperative mode. The difference of course, is that the structure of such a complex acronym as *grep* is not built from a linear association of phonemes as would any simple lexical unit like *car*, *table*, *wall*, *tree*, etc. but from a complex acronym the initials of which represent only part of the sentence it is meant to stand for.

The additional layer of meaning requires knowledge of the individual constituents of the componym which are called Minimal Informational Cooperative Units. All componyms are made up from MICUs. The componyms may later undergo a process of inflection or of affixation. Other examples of componyms capable to represent complex semiotic objects are: *To laserize*, which results from the inflection of the acronym laser (*Light Amplification by Stimulated Emulsion of Radiation*). Once inflected, it behaves as an ordinary verb. *ASCIIbetical order*, is an affixed complex componym built from the acronym ASCII (*American Standard Code for Information Interchange*). As clearly stated above, componyms consist in MICUs, instead of being the result of the combination of phonemes that compose ordinary words. Like acronyms, componyms are built from initials

of words, but unlike acronyms, they undergo a process of derivation or of inflection which fallaciously make them appear as ordinary words.

In effect, comonyms are written with lowercase characters, and this adds to their confusion with simple lexical units. In association with hypertexts and hyperlinks, comonyms shape an innovative syntax able to build complex meanings beyond the limits of the canonical grammatical sentence, thus opening new paths in network thinking, writing and reading. Other examples of coinages found in cyberspace such as *CU B4 noon* are even more striking. Here the sentence *See you before noon* is reduced to two initials of consecutively a verb and a pronoun: C for the verb *to see* and U for the pronoun *you*. These two initials are associated with a hybrid form composed of the first syllable of the preposition *before* and the alphanumeric character 4 to obtain B4. Eventually, the noun *noon* or another substantive is added to form a complete sentence. Because MICUs permit to code units of information instead of phonemes, they allow for a much wider scope to construct complex meaning and transmit it in a like manner. Cellular phone SMS writers make an extensive use of this new type of writing which is not truly writing.

What misses for the moment, are the conventional rules to code and decode the new syntax to ensure a practical and reliable grammar. This is precisely the new linguistic field for linguists to investigate and explore with the view to elaborate for each language new syntactic rules that would account for the integration of comonyms into their specific grammars. The new rules ought to account closely both for the new types of relationships between the various units and structures which make up the new type of discourse, and for the multimodality incorporated into the new texts initiated by electronic writing, so distinct from those induced by the printed text.

The novel linguistic tools able to coin complex neologies may definitely free discourse from the tenets of linguistic closure. In a sense, with comonyms, the word looks more like a complex interface. The syntagmatic relationships are cyclically woven, depending on the paradigmatic demands of the situation of communication. With regular practice, after being arranged and identified as such, the comonyms acquire a socially acknowledged acceptability and stability, which permit them to integrate the 'trésor de la langue', and be considered as ordinary simple lexical units. In other words, once internalized a comonym like *laser*, loses its status as a complex acronym or comonym, and behaves as if it were a simple word but loaded with several layers of subsumed meaning. Eventually, only specialists of

language may be interested in their etymological structure. After all, it is wise to consider Bolter's pronouncement¹¹ that 'the best way to understand electronic writing today is to see it as the remediation of printed text, with its claim to refashioning the presentation and status of alphabetic writing itself.' Therefore, one may feel entitled to consider with Ferris¹² that 'as communication theorists, we should actively work to understand concepts inherent in traditional writing, and as actively work to shape the development of electronic writing.'

2 CONCLUSION

A serious impetus has already been given to this perspective with the consecration of hypertext by the internet as a starting structure towards the formation of novel experiences of conceiving, storing, and disseminating knowledge. The new writing space has considerably updated reflection upon the notion of text as a cognitive interface between a producer and consumer of knowledge, between author and readership. This new expression modality has definitely windswept the principle of 'text closure' already undermined by 'Barthes'¹³ distinction between readerly and writerly texts, and has propelled discourse towards an ever extending externality. New reading processes need now to be devised accordingly to meet with the new literacy demands. The road to linguistic innovation is wide open to cater for the new information overload.

Our stance is that despite their timid emergence, comonyms have the necessary potential to rapidly compose a new complex grammar. Their hypertextual nature is an added value for their inclusion in network writing. The process has already started with hypertext and hypermedia, but the new grammar lacks a fundamental structure which lexicogenic processes like comonyms are largely fit to assume. Once the need for these structures is clearly expressed, which is precisely the case in the present paper, what remains is the identification of the potential relationships that may pragmatically render it possible to compose a new grammar.

¹¹Bolter, Jay, David. *Writing Space: Computers, Hypertext, and the Remediation of Print*, Routledge, 2001. P. 26.

¹²Shamila Pixy Ferris. *Writing Electronically: The Effects of Computers on Traditional Writing*.

¹³See R. Barthes, "S/Z", Blackwell Publishing, 1990. P. 15.

REFERENCES

- [1] Barthes, R. *S/Z*, First published in the United Kingdom by Blackwell Publishing Ltd, 1990.
- [2] Bolter, J.D. *Writing Space: Computers, Hypertext, and the Remediation of Print*, Hillsdale, NJ: Erlbaum, 2001.
- [3] Danesi, M. *Understanding Digital Semiotics*, OUP New York. 2002.
- [4] Eisenstein, E.L. *The Printing Press as an Agent of Change*, Vol 1 & 2, CUP 1979.
- [5] Eisenstein, E.L. *The printing revolution in early modern Europe*. Cambridge, UK: Cambridge University Press. 1983.
- [6] Ferris, S.P. Writing Electronically: The Effects of Computers on Traditional Writing <http://quod.lib.umich.edu/jjep/3336451.0008.104?view=text;rgn=main>.
- [7] Fodil M.S. *MICUs*, “Componyms and the Triple Articulation of Cyber Language”, El Khitab, Vol.7, June 2010.
- [8] Kramer S.N. *History Begins in Sumer*, The University of Chicago Press, 1963.
- [9] Logan, R.K. *The extended mind model of the origin of language and culture*. In: N. Gontier, J.P. Van Bendegem and D. Aerts (Eds.) Evolutionary epistemology, language and culture. Dordrecht. Springer, 2006.
- [10] Mc Luhan, M. *Understanding Media: The Extensions of Man*, MIT Press Edition, 1994.
- [11] Martinet, A. *Eléments de linguistique générale*. Armand Colin, 1998.
- [12] Ong, W.J. *Orality and Literacy: The Technologizing of the World*, Routledge, 2002.
- [13] Saussure F. *Le Cours de Linguistique Générale*, ENAG, 1994.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

MVDR beamformer model for array response vector mismatch reduction

Suhail Najm Shahab & Ayib Rosdi Zainun

Faculty of Electrical and Electronics Engineering, Universiti Malaysia Pahang, Pahang, Malaysia

Essa Ibrahim Essa

Department of Computer Science, College of Computer Science and Mathematics, Tikrit University, Tikrit, Iraq

Nurul Hazlina Noordin & Izzeldin Ibrahim Mohamed

Faculty of Electrical and Electronics Engineering, Universiti Malaysia Pahang, Pahang, Malaysia

A. Omar Khaldoun

School of Computer and Communication Engineering, University Malaysia Perlis, Perlis, Malaysia

ABSTRACT: Beamforming algorithms attempt to extract a desired User-Of-Interest (UOI) from the background noise and interfering signals. The performance of the beamforming algorithm is evaluated based on various QoS criteria such as beampattern accuracy and Signal-to-Interference-plus-Noise-Ratio (SINR). In this paper, the null-forming constrain is added to the single linear constrain of Minimum Variance Distortionless Response (MVDR) to overcome the effect of finite snapshots problem and the array response vector imprecision. This constraint addition improves the null-forming at the User-Not-Of-Interest (UNOI) direction. This work presents a new approach for extract the accurate array response vector. Numerical results show the robustness of the proposed approach to alleviating finite data snapshots effect. Moreover, this technique minimizes the sidelobe level, accurate beam shape to the UOI direction and pattern null in the UNOIs directions.

1 INTRODUCTION

The growth in the number of wireless devices and applications has led to a crowding of the wireless spectrum and more stringent requirements for receiver designs. Radio frequency interference continues to be a persistent problem in many communication systems and will potentially exacerbate as the unused wireless spectrum continues to shrink. There are, in general, two types of interfering signals; (1) Intentional jammers used in military applications, such as Electronic Warfare (EW). (2) Unintentional, yet harmful interference, primarily associated with wireless commercial systems (Van Trees, 2002). In recent years, there has been a rapid increase in the number of wireless devices for both commercial and defense applications. This has been adding strain on the spectrum utilization of wireless communication systems. Because there exists a limited amount of available frequency spectrum, interference is bound to occur as the spectrum saturates (Gross, 2011). Antenna arrays are used in wireless communications to focus electromagnetic energy on a signal of interest while simultaneously minimizing energy in jammer directions.

The Minimum Variance Distortionless Response (MVDR) or Capon beamformer (Capon, 1969) is one of the adaptive optimum statistical beamformers which assures a distortionless response for a predefined steering direction (Van Trees, 2002, Gross, 2015, Godara, 1997). The basic idea of the MVDR technique is to estimate the beamforming excitation coefficients in an adaptive way by minimizing the variance of the residual interference and noise whilst enforcing a set of linear constraints to ensure that the real user signal is not distorted (Pan et al., 2014).

The most common MVDR problem is that the signal model must be quite accurate in order not to form unity gain in the UOI direction nulls in the direction of the UNOI. There are many ways to make the MVDR beamformer robust against this error such as diagonal loading (Gu et al., 2008, Lin et al., Feb. 2007), beamspace processing (Feldman and Griffiths, 1994), spatial averaging (Van Trees, 2002). Another problem with the MVDR beamformer is the finite size of data snapshots (Van Trees, 2002, Wax and Anu, 1996a, Fertig, 2000, Mestre and Lagunas, 2006, Chen and Lee, 2012, Ghadian et al., 2015) and the array response

vector uncertainty (Wax and Anu, 1996b, Besson and Vincent, 2005). When the size of data snapshots is small will result in a poorly represented beampattern and degrades the MVDR performance.

Many attempts have been made (Lin et al., Feb. 2007, Gu et al., 2008) in order to find the optimal weight vector. For example, the idea of diagonal loading (Lin et al., Feb. 2007, Gu et al., 2008) is to adapt a covariance matrix by adding a displacement value to the diagonal elements of the estimated covariance matrix. The study carried out by (Lu et al., 2013), the authors demonstrate the array response vector mismatch due to array calibration misadjustment based on MVDR technique. The method used an iterative algorithm to reconstructing the covariance matrix to overcome the array response pointing errors. More recently, a study carried out by (Abdulrahman et al., 2015) for enhancing the MVDR performance against the array response error by replacing the reference element in the linear antenna array to be in the middle. The results show the enhanced model the minimum number of data snapshots required to produce satisfactory resolution is 30 snapshots.

However, the effects of finite snapshot size on the beampattern accuracy and the output SINR are still unknown from the expressions. The present work introduces a new method to estimate the array response vector by adding the null-forming linear constraint to the MVDR technique. Therefore, the weights coefficients calculated to place null toward the UNOI direction accurately, and unity gain response toward the direction of user-of-interest. Simulation results confirm the accuracy of the theoretical results.

The remainder of this paper is organized as follows. In section 2, MVDR beamformer based on linear antenna array design method with the signal propagation model is described. The simulation results and performance evaluation are provided in Section 3. Finally, in Section 4, the paper's conclusions and summary of MVDR performance are described.

2 MVDR BEAMFORMER DESIGN MODEL

The basic theory of the beamforming algorithm and the signal structure is presented in this section. The signal model considers L signals impinging on a ULA of M isotropic antenna elements, and the spacing between adjacent antennas is a half of wavelength. Assume that L signal coming from angles of θ_i and ϕ_i is incident upon an antenna array of M elements, shown in Figure 1. Here, the impinging angles of θ and ϕ are the azimuthal and elevation angles, respectively.

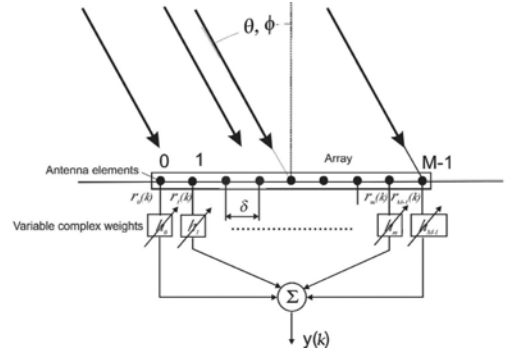


Figure 1. Uniform Linear antenna array geometry.

The received signal, $r_m(k) \in \mathbb{C}^{M \times M}$, at the m^{th} antenna at the k^{th} snapshot incident upon the antenna array can be written as:

$$r_m(k) = \sum_{d=1}^D x_d(k) v_d(\theta, \phi) + \sum_{i=1}^L x_i(k) v_i(\theta, \phi) + \sum_{m=1}^M n_m(k) \quad (1)$$

where $x_d(k)$, $x_i(k)$, and $n_m(k)$ denote the d^{th} user-of-interest signals, i^{th} interference signals and additive background White Gaussian noise at the m^{th} elements, respectively. Among those L incident signals, it is assumed that $x_d(k)$ is the desired user-of-interest and $x_i(k) + n_m(k)$ are the user-not-of-interest signals. The array response vector, $v(\theta, \phi)$ is a $\mathbb{C}^{1 \times M}$ of a Uniform Linear Array (ULA) with M antenna elements where (θ, ϕ) are the DOAs of the l^{th} signal component given as (Godara, 2004, Godara, 1997):

$$v(\theta, \phi) = [1, e^{-j\beta\delta\sin\theta\sin\phi}, \dots, e^{-j(M-1)\beta\delta\sin\theta\sin\phi}]^* \quad (2)$$

where $\beta = 2\pi/\lambda$ is the free-space wavenumber, δ is the spacing between adjacent antenna elements and λ is the free-space wavelength. The $\theta \in [-\pi/2, \pi/2]$, $\phi \in [0, \pi/2]$ and $(.)^*$ denote the complex conjugate. The $v(\theta, \phi)$ is a function of the incident angles, the location of the antenna, and the array geometry. It plays an important role in smart antenna systems, containing information of the impinging angles. The output of the beamformer at the k^{th} snapshots, $y(k)$ after signal processing is defined as:

$$y(k) = \sum_{m=1}^M h_m^* r_m(k) \quad (3)$$

where h is a complex multiplicative weight vector given as $[h_0, h_1, \dots, h_m, h_{M-1}]^T$ multiplied by the received signal at the m^{th} antenna element and

$(\cdot)^*$, $(\cdot)^T$ denotes respectively the complex conjugate transpose of a vector or matrix and transpose of a vector or matrix. The array cross-correlation (covariance) matrix $\Gamma_r \in \mathbb{C}^{M \times M}$ matrix, is defined as (Krim and Viberg, 1996):

$$\Gamma_r = \{r_m(k)r_m^*(k)\} \quad (4)$$

The array covariance matrix Γ_r in Eq. (4) is the statistical second-order property of the impinging signals. In real applications, Γ_r is estimated using the received array snapshots. The estimated array covariance matrix is given by (Gross, 2015):

$$\hat{\Gamma}_r \cong \Gamma_d + \Gamma_{i+n} \cong \frac{1}{K} \sum_{k=1}^K r_m(k)r_m^*(k) \quad (5)$$

$$\Gamma_d = \sum_{d=1}^D \sigma_d^2 v_d(\theta, \phi) v_d^*(\theta, \phi) \quad (6)$$

$$\Gamma_{i+n} = \sum_{i=1}^I \sigma_i^2 v_i(\theta, \phi) v_i^*(\theta, \phi) + \sigma_n^2 \Lambda_m \quad (7)$$

where K is the number of available snapshots. Γ_d denotes the array correlation matrix corresponding to the desired user-of-interest and Γ_{i+n} refer to the array correlation matrix corresponding to the undesired user-not-of-interest. The terms σ_d^2 , σ_i^2 and σ_n^2 denotes the real user, interference, and noise powers. $\Lambda_m \in \mathbb{R}^{M \times M}$ stands for the identity matrix. It is known from the literature that the optimization criterion for MVDR (Capon, 1969) forms weights in a way that will attempt to maintain unity gain of the beamformer in the beam angle direction while steering nulls in the direction of interference (Souden et al., 2010). The weights are calculated by solving the following minimization equations with unity gain restraint:

$$h_{MVDR} = \arg \min_{0 \leq h \leq 1} E \{ |y(k)|^2 \} \quad (8)$$

$$\min_h h^* \hat{\Gamma}_r h \text{ s.t. } h^* v_d(\theta, \phi) = 1 \quad (9)$$

The above equations are solved by using Lagrange multipliers and the MVDR weight (h_{MVDR}) is given as (Balanis and Ioannides, 2007):

$$h_{MVDR} = \frac{v(\theta, \phi) \hat{\Gamma}_r^{-1}}{v^*(\theta, \phi) \hat{\Gamma}_r^{-1} v(\theta, \phi)} \quad (10)$$

Zero-Forcing (ZF) beamforming technique (Davies, 1967) have been used extensively for interference suppression purpose in wireless communication. ZF method able to cancels several plane waves impinging from known directions and directs the mainbeam toward the desired user-of-interest source. Assume that the desired signal with

steering vector $v_d(\theta, \phi)$ and I interference sources are impinging on the array. The null-forming weight vector is calculated based on the unity power reception in the direction of the desired signal and null-forming is used to steer the nulls (zero or near-zero antenna power) reception in the interference sources directions with steering vectors $v_i(\theta, \phi)$ is given by (Qamar and Khan, 2009, Friedlander and Porat, 1989):

$$h^* v_d(\theta, \phi) = 1 \quad (11)$$

$$h^* v_i(\theta, \phi) = 0; \quad i = 2, 3, \dots, M \quad (12)$$

However, the computational burden associated with this approach is quite high and makes it difficult for real-world applications. Based on the idea of ZF method of two steering vector. The proposed array response vector to overcome the limited number of data snapshots to obtain high resolution in term of beampattern accuracy can be defined as:

$$v^p(\theta, \phi) = [\Lambda_m - v_i(\theta, \phi) \{v_i^*(\theta, \phi) v_i(\theta, \phi)\} v_i^*(\theta, \phi)] v_d(\theta, \phi) \quad (13)$$

Then, the ZF algorithm and MVDR technique can be combined to achieve the (beam+null)-forming. The combined constraints can be rewritten as:

$$\min_h h^* \hat{\Gamma}_r h \text{ s.t. } h^* v_d(\theta, \phi) = 1; \quad h^* v_i(\theta, \phi) = 0 \quad (14)$$

Thus, the new complex weight vector calculation according to the proposed array response vector from Eq (13):

$$h_{MVDR}^p = \frac{v^p(\theta, \phi) \hat{\Gamma}_r^{-1}}{v^{p*}(\theta, \phi) \hat{\Gamma}_r^{-1} v^p(\theta, \phi)} \quad (15)$$

The null-forming beamformer can be formulated as:

$$h_{MVDR}^p V^p(\theta, \phi) = h_{MVDR}^p [v_d^p(\theta, \phi), v_i^p(\theta, \phi), \dots, v_{M-1}^p(\theta, \phi)] \quad (16)$$

where $V^p(\theta, \phi)$ is the array response matrix which containing the desired array response vector and the interfering signals response vector.

Antenna radiation patterns are typically expressed in terms of radiated power. The output power is defined as (Godara, 1997):

$$P_y = E \{ y(k) y^*(k) \} = h^* E \{ y(k) y(k) \} h = h^* \hat{\Gamma}_r h \quad (17)$$

Equation (16) can be rewritten as:

$$P_y = h^\dagger \Gamma_d h + h^\dagger \Gamma_{i+n} h = P_d + P_{i+n} \quad (18)$$

$$P_d = \sum_{d=1}^D \sigma_d^2 |h^\dagger v_d(\theta, \phi)| \quad (19)$$

$$P_{i+n} = \sum_{i=1}^I \sigma_i^2 |h^\dagger v_i(\theta, \phi)| + \sigma_n^2 \quad (20)$$

where the P_d denote the power of the desired signal and P_{i+n} refer to the power output in the direction of UNOI. Finally, the SINR is defined as the ratio of the average power of the desired signal divided by the average power of the undesired signal computed as (El Zooghby, 2005):

$$SINR \triangleq \frac{P_d}{P_{i+n}} \triangleq \frac{\sum_{d=1}^D \sigma_d^2 |h^\dagger v_d(\theta, \phi)|}{\sum_{i=1}^I \sigma_i^2 |h^\dagger v_i(\theta, \phi)| + \sigma_n^2} \quad (21)$$

3 NUMERICAL RESULTS AND ANALYSIS

In this section, the results of this two beamformer are discussed where Matlab® platform has been used to model the performance results in terms of mathematical functions. To compare the MVDR algorithm and the proposed approach, we perform a simulation according to the parameters as given in Table 1. It has been assumed that all users are stationary in a multipath fading environment and the performance of each algorithm is evaluated under the same noise and interference conditions.

3.1 Case I

First, the performance of the conventional MVDR beamformer and the proposed method are investigated and the weight vector calculated based on Eqs. 10 and 15. The first case simulation show the beampattern for a different number of snapshots

Table 1. Key system parameters.

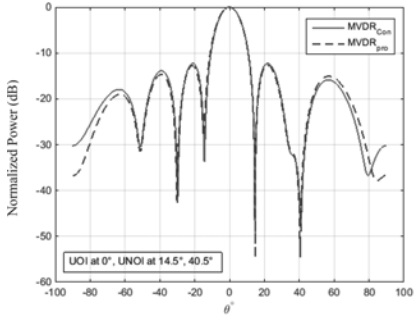
Key system parameters	Values
Array antenna configuration	LAA
Antenna type	Isotropic
Carrier frequency (F_c)	2.6 GHz
Beam scanning range	$\pm 90^\circ$ (Azimuth)
Number of element (M)	8
Element spacing (δ)	$\lambda/2$
# UOI	1
# UNOIs	2
SNR [dB]	10
INR [dB]	10
Snapshots (K)	500, 100, 50, 10

of the intended user and two interfering sources are introduced from two different azimuth angles 14.5° and 40.5° and the desired information signal arrives at the user from an azimuth angle of 0° . The antenna array is a ULA and the antenna elements are separated by 0.5λ and the noise level and the interference level is fixed of 10 dB. The element spacing, δ is set to 0.0577 m to satisfy the half of the wavelength separation between neighboring elements. Different numbers of data samples are used and the results are shown in Figs. 2–5 to illustrate a comparison between these two algorithms.

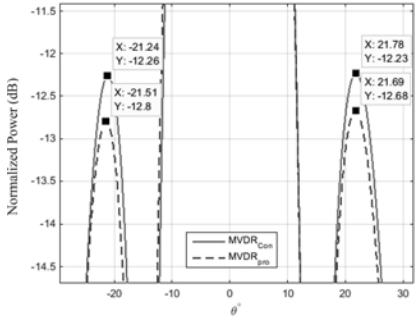
Figure 2(a) illustrate a comparison between the conventional MVDR (MVDR_{con}) and the proposed MVDR (MVDR_{pro}) algorithms for 500th snapshots. It is evident that the MVDR_{pro} algorithm is more efficient, all UNOIs source are perfectly null with more precision and hence, the peaks of SLLs are lower than MVDR_{con} as it depicted in Figure 2(b)-(c). It is clearly shown from Figure 2(b) that the SLL for the MVDR_{pro} found to be -12.6 dB compare to -12.2 dB for the MVDR_{con}. It shown the SINR at the output of the MVDR_{pro} beamformer is 21.6 dB whilst the MVDR_{pro} of 18.8 dB. It can observe that the beampattern are fairly similar for both algorithms.

Secondly, the arriving signals are time varying, therefore; the calculations are based on time snapshots of the incoming signal. Figures 3(a)-(c) show the beampattern for $K = 100$ th. Figure 3(a) shows that the direction of the mainlobe for both algorithms is steered towards the desired direction, 0° . It also shows that the height of SLL from MVDR_{con} is higher than that from MVDR_{pro} by 0.8 dB, therefore, the interference suppression from MVDR_{pro} is less than from MVDR_{con} as reveals from Figure 3(b). The graph shows that the MVDR_{pro} achieves null for $\theta = 14.5^\circ$ and 40.5° of -52 dB and -41 dB comparing -33 dB and -36 dB to achieved by MVDR_{con}. The SINR obtained for MVDR_{con} and MVDR_{pro} are 18.2 dB, and 20.3 dB, respectively. That means MVDR_{pro} achieving deep null even the snapshots size of 100.

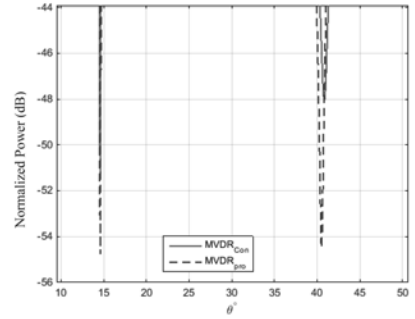
Thirdly, the simulation that illustrates the 50th data snapshots is shown in Figure 4(a)–(c). The beam angle was steered toward 0° , and MVDR_{pro} clearly present null at the two interfering signals whereas the null position of MVDR_{con} shifted by 3.5° at 40.5° as illustrated in Figure 4(c). Figure 4(c) shows that the interference suppression of MVDR_{pro} is lower than MVDR_{con} due to the higher height of SLLs. As can be seen from Figure 4(c), the UNOIs signals that arrive from 14.5° & 40.5° are rejected by the MVDR_{con} and the MVDR_{pro} beamformer creates a null of -42 dB, -32 dB and -49 dB, -37 dB, respectively. To illustrate the SINR performance of the proposed approach, SINR value is ≈ 17.5 dB;



(a)



(b)

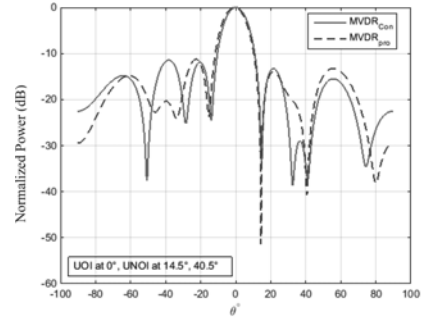


(c)

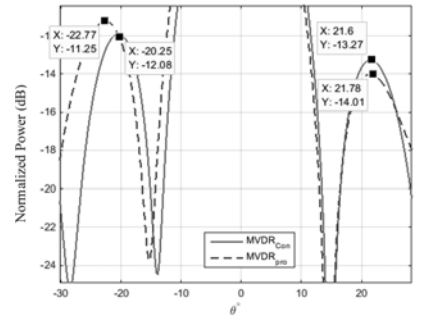
Figure 2. MVDR beamformer; UOI at 0° , UNOIs at 14.5° , 40.5° , $M = 8$, $K = 500$; (a) typical MVDR beam-pattern. (b) Zoom in SLLs pattern. (c) Zoon in the null-forming pattern.

this represents a 9% improvement to the $MVDR_{con}$ of 16.1 dB.

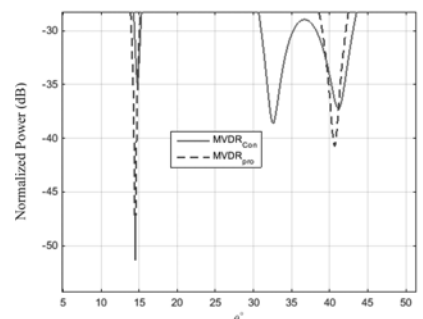
Lastly, the graphical representation of a beam-pattern is created by plotting P_y versus all possible incident directions. The number of snapshots (K) creating $\hat{\Gamma}_r$ is 10 as demonstrates in Figure 5(a)–(e). Figure 5(a) is an example of a normalized beam-pattern when a ULA with eight antennas is used, the spacing between adjacent antennas is half of the carrier frequency, the desired user-of-interest comes from 0° , two interfering signals come from 14.5° and 40.5° . It can be seen that there is a sin-



(a)



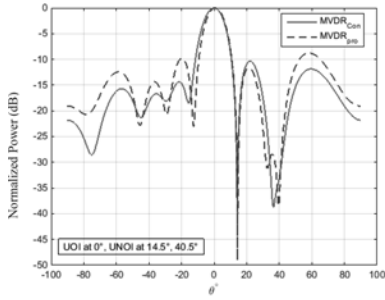
(b)



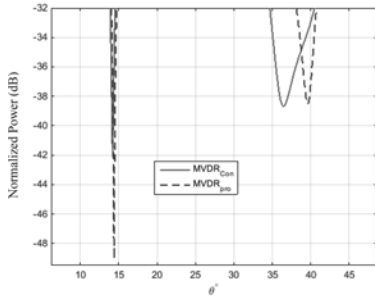
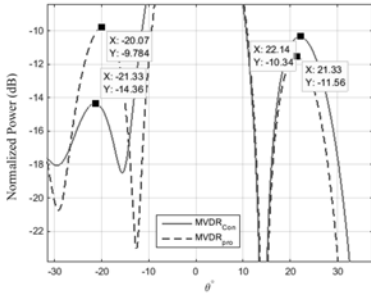
(c)

Figure 3. MVDR beamformer; UOI at 0° , UNOIs at 14.5° , 40.5° , $M = 8$, $K = 100$; (a) typical MVDR beam-pattern. (b) Zoom in SLLs pattern. (c) Zoon in the null-forming pattern.

gle dominant mainlobe peak directed toward 0° . The $MVDR_{pro}$ mainbeam are steered toward the incident angle of the desired user-of-interest while the $MVDR_{con}$ provides the mainbeam steered 10° off the desired target direction. From Figure 5(b), the height of the SLL found to be -2.8 dB and -4.4 dB for $MVDR_{pro}$ and $MVDR_{con}$, respectively. Figure 5(c) shows that the null-forming of the interfering signal from 14.5° and 40.5° is nullified by -33 dB and -23 dB, while in comparison with $MVDR_{con}$ the interfering signal from 14.5° is suppressed by -24 dB and -23 .



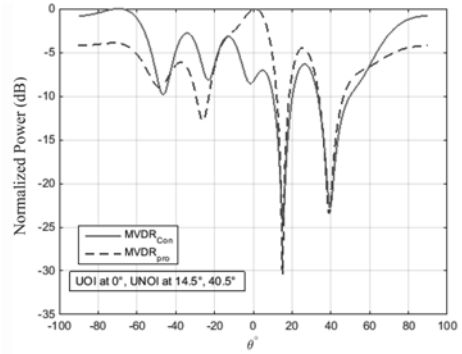
(a)



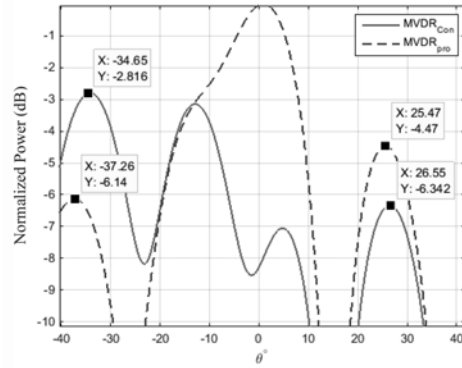
(c)

Figure 4. MVDR beamformer; UOI at 0° , UNOIs at 14.5° , 40.5° , $M = 8$, $K = 50$; (a) typical MVDR beam-pattern. (b) Zoom in SLLs pattern. (c) Zoon in the null-forming pattern.

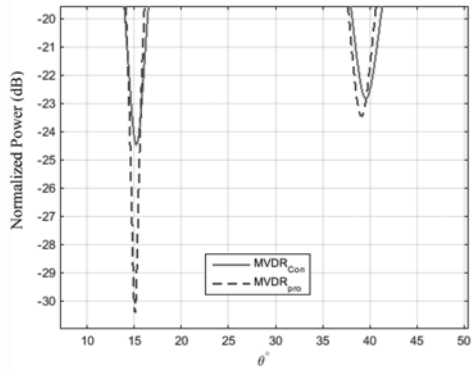
Figures 5(e)-(f) show 3D beampattern for azimuthal and elevation scan angles plots of the $MVDR_{con}$ and $MVDR_{pro}$. The power is measured in dB and a color bar is used for a sense of the relative scale of the power. The inner rectangle dashed black line represents the null width that encompasses the UNOIs target while the main-lobe represents by 'Mainlobe' in the both figures. Furthermore, it can be easily seen by comparing these two figures, the null width in the θ° by $MVDR_{pro}$ narrower than $MVDR_{con}$ and the null-forming deeper than the conventional MVDR. It is observed that the SINRs of the $MVDR_{con}$ is -11.4 dB, whereas the $MVDR_{pro}$ results show 15.1 dB giving a 33% improvement.



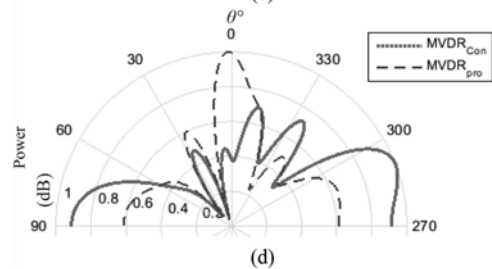
(a)



(b)



(c)



(d)

Figure 5. (Continued).

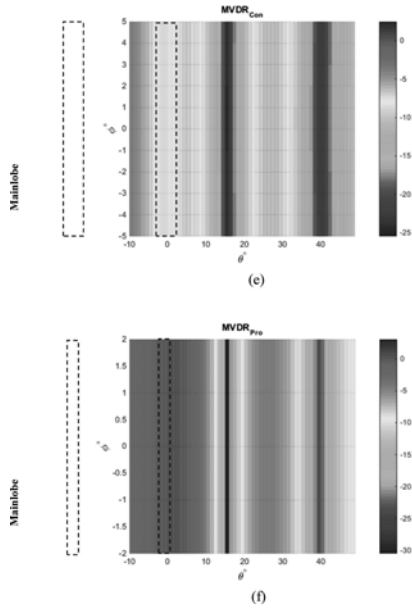


Figure 5. MVDR beamformer; UOI at 0° , UNOIs at 14.5° , 40.5° , $M = 8$, $K = 10$; (a) typical MVDR beam-pattern. (b) Zoom in SLLs pattern. (c) Zoon in the null-forming pattern. (d) Polar power pattern. (e)-(f) 3D beampattern in term of azimuth and elevation angles.

3.2 Case 2

The simulation in case two divided into two scenarios, the first scenario illustrative comparison between the performance of the proposed method and the conventional MVDR. Figure 6 present the SINR obtained for the range of the 1st snapshot to the 100th snapshots. The solid red line represents the conventional ($MVDR_{con}$) and the dashed blue line demonstrates the ($MVDR_{pro}$) as illustrate in Figure 6.

The result shows that $MVDR_{pro}$ has better tracking ability to compute the final weight vector compared to $MVDR_{con}$ when the data snapshots less than 50. $MVDR_{pro}$ show superior performance, for example, at $K = 30$, it found that the SINR for $MVDR_{con}$ is equal to 12.1 dB compared 16.7 dB obtained by $MVDR_{pro}$. The SINR for $MVDR_{con}$ increase gradulay after the snapshots size reach more than 50. It is observed that the SINRs of the $MVDR_{con}$ and proposed algorithm increase with the increase of the K . The output SINR for the proposed method giving a 56% improvement for data snapshots size ≤ 100 . Since the weight coefficients depend on the covariance matrix estimation and the array steering response. It enhances the output SINR by placing accurate nulls in the direction of interferences. Therefore, the $MVDR_{pro}$ seems to have stable performance.

Figure 7 displays the SINR vs SNR is varied from -20 to 20 for both beamformers. The performance of both algorithms evaluated using eight linear antenna elements used ($M = 8$), the spacing between antennas is half of the carrier wavelength for fixed training data size $K = 150$. There are 4 signals impinging from 10° , $\pm 25^\circ$ and -60.7° . The signal arriving from 10° is the desired user-of-interest and all others are interfering signals. The solid red line represents the conventional $MVDR_{con}$ and the dashed blue line demonstrates the $MVDR_{pro}$ as illustrate in Figure 7. This increase in resolution of $MVDR_{pro}$ is due to the interference signal being perfectly nulled even of at a range of SNR.

It clearly is observed in this figure that the $MVDR_{con}$ beamformer does increase the resolution due to the data did not actually estimate the location exactly. Furthermore, it can be seen that the proposed approach has accurate beampattern toward the mainlobe target and the null-forming toward the disturbance sources. Therefore, the $MVDR_{pro}$ algorithm seems to have stable and good performance regardless of the SNR values. In addition, the mean SINR achieved by the $MVDR_{pro}$ is always greater than the SINR achieved by the MVDR technique, and their difference increases with increasing SNR.

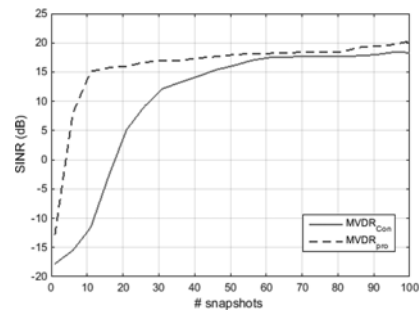


Figure 6. Output SINR versus K for array response mismatch.

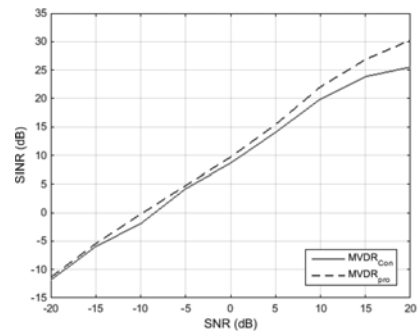


Figure 7. Output SINR versus SNR for array response mismatch.

4 CONCLUSION

This paper looked at the MVDR beamforming problem in a multiple access interference environment. MVDR beamformer is sensitive to errors such as array response mismatch and the effect of finite snapshots. To keep the snapshot requirement to a minimum, the proposed method provide a robust solution for MVDR array response vector mismatch problem and limited data snapshots. New beamforming method is developed for uniform linear array configuration. This method has high-resolution and is statistically efficient and consistent. The performances are evaluated and compared MVDR technique. The present method provides different performances between in terms of beam-pattern accuracy and SINR improvement. From the results of the simulation, the conclusions drawn are that the proposed beamformer successfully increased the MVDR resolution and accurately null the user-not-of-interest signals. The mainlobe steer to the user-of-interest direction even with very small data snapshots. In general, we can observe that for these two methods, the beampattern is almost the same for a large number of data snapshots.

ACKNOWLEDGMENT

This research was supported by Universiti Malaysia Pahang, through the Fundamental Research Grant Scheme (FRGS) funded by Ministry of Education (RDU 140129).

REFERENCES

- Abdulrahman, O. K., Rahman, M. M., Hassnawi, L. & Ahmad, R. B. 2015. Modifying MVDR Beamformer for Reducing Direction-of-Arrival Estimation Mismatch. *Arabian Journal for Science and Engineering*, 1–14.
- Balanis, C. A. & Ioannides, P. I. 2007. *Introduction to smart antennas*, Arizona, USA, Morgan and Claypool Publishers.
- Besson, O. & Vincent, F. E. 2005. Performance analysis of beamformers using generalized loading of the covariance matrix in the presence of random steering vector errors. *Signal Processing, IEEE Transactions on*, 53, 452–459.
- Capon, J. 1969. High-resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE*, 57, 1408–1418.
- Chen, Y. L. & Lee, J.-H. 2012. Finite data performance analysis of LCMV antenna array beamformers with and without signal blocking. *Progress In Electromagnetics Research*, 130, 281–317.
- Davies, D. 1967. Independent angular steering of each zero of the directional pattern for a linear array. *Antennas and Propagation, IEEE Transactions on*, 15, 296–298.
- El Zooghby, A. 2005. *Smart antenna engineering*, Norwood, MA, USA, Artech House, Inc.
- Feldman, D.D. & Griffiths, L. J. 1994. A projection approach for robust adaptive beamforming. *Signal Processing, IEEE Transactions on*, 42, 867–876.
- Fertig, L.B. Statistical performance of the MVDR beamformer in the presence of diagonal loading. *Sensor Array and Multichannel Signal Processing Workshop*. 2000. Proceedings of the 2000 IEEE, 2000. IEEE, 77–81.
- Friedlander, B. & Porat, B. 1989. Performance analysis of a null-steering algorithm based on direction-of-arrival estimation. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37, 461–466.
- Ghadian, M., Jabbarian-Jahromi, M. & Kahaei, M. 2015. Recursive Sparsity-based MVDR Algorithm for Interference Cancellation in Sensor Arrays. *IETE Journal of Research*, 1–9.
- Godara, L. C. 1997. Application of antenna arrays to mobile communications. II. Beam-forming and direction-of-arrival considerations. *Proceedings of the IEEE*, 85, 1195–1245.
- Godara, L. C. 2004. *Smart antennas*, Boca Raton, CRC press.
- Gross, F. 2015. *Smart antennas with matlab: principles and applications in wireless communication*, McGraw-Hill Professional.
- Gross, F.B. 2011. *Frontiers in antennas: next generation design & engineering*, McGraw-Hill New York, NY, USA.
- Gu, Y. J., Shi, Z.-G., Chen, K. S. & Li, Y. 2008. Robust adaptive beamforming for steering vector uncertainties based on equivalent DOAs method. *Progress In Electromagnetics Research*, 79, 277–290.
- Krim, H. & Viberg, M. 1996. Two decades of array signal processing research: the parametric approach. *Signal Processing Magazine, IEEE*, 13, 67–94.
- Lin, J.-R., Peng, Q.-C. & Shao, H.-Z. Feb. 2007. On diagonal loading for robust adaptive beamforming based on worst-case performance optimization. *ETRI journal*, 29, 50–58.
- Lu, Z., Li, Y., Gao, M.-L. & Zhang, Y. 2013. Interference covariance matrix reconstruction via steering vectors estimation for robust adaptive beamforming. *Electronics Letters*, 49, 1373–1374.
- Mestre, X. & Lagunas, M. Á. 2006. Finite sample size effect on minimum variance beamformers: Optimum diagonal loading factor for large arrays. *Signal Processing, IEEE Transactions on*, 54, 69–82.
- Pan, C., Chen, J. & Benesty, J. 2014. Performance study of the MVDR beamformer as a function of the source incidence angle. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22, 67–79.
- Qamar, R. A. & Khan, N. M. Null steering, a comparative analysis. *IEEE 13th International Multitopic Conference (INMIC'09)*, 2009. IEEE, 1–5.
- Souden, M., Benesty, J. & Affes, S. 2010. A study of the LCMV and MVDR noise reduction filters. *IEEE Transactions on Signal Processing*, 58, 4925–4935.
- Van Trees, H. L. 2002. *Optimum array processing: part IV of detection, estimation, and modulation theory*, New York, Wiley.
- Wax, M. & Anu, Y. 1996a. Performance analysis of the minimum variance beamformer. *Signal Processing, IEEE Transactions on*, 44, 928–937.
- Wax, M. & Anu, Y. 1996b. Performance analysis of the minimum variance beamformer in the presence of steering vector errors. *Signal Processing, IEEE Transactions on*, 44, 938–947.

Enhancing quality of service by balancing the traffic load in Mobile Ad hoc Networks (*MANETs*)

Guidoum Amina & Aoued Boukelif

Department of Electronics, Djillali Liabes University, Sidi Bel Abbes, Algeria

ABSTRACT: Network congestion is the main reason for packet loss, longer delay in streaming multimedia applications. AODV (Ad hoc on-demand Distance Vector) routing protocol is one of the well-known and efficient on-demand MANET (Mobile Ad-Hoc Networks) protocols. AODV currently does not support Quality of Service (QoS) and has no load balancing mechanism. We present a new solution to achieve better load balancing by introducing, a new procedure to avoid congested nodes and the integration of the concept of traffic load in route discovery process and connectivity management in this way the least congested path will be chosen by the source node. The performed simulation and comparison between the AODV protocol and AODV-balanced show that this last can improve the QoS in Mobile ad hoc networks under different conditions using NS2 simulator.

Keywords: connectivity management, MANET, NS2, QOS, route discovery

1 INTRODUCTION

Mobile Ad-Hoc Networks (MANETs) are wireless networks where a collection of mobile nodes can dynamically vary the topological structure. With respect to the more widely used mobile cellular networks (MANETs do not use any form of fixed infrastructure or centralized administration). A set of ad hoc routing protocols has been proposed in the IETF's MANET group to ensure the network connectivity [1]. They operate in either proactive or reactive modes. Ad hoc network presents many specific problems which had influence on solution that assure QoS. The level of Service that a user obtains from a network is known as the Quality of Service. The goal of QoS offered is to ensure a better delivery of information carried by the network, and a better utilization of the network's resources. The network provides a set of service guarantees such as minimum bandwidth, maximum delay, and maximum loss rate while transporting a packet stream from the source to the destination [2]. the routing protocols Ad hoc On-demand Distance Vector (AODV) [3] adopts a purely reactive strategy: it sets up a route on demand at the start of a communication session, and uses it till it breaks, after which a new route setup is initiated. It is based on the principle of distance vector routing. Given its characteristics, this protocol has become widely known and has been a lot of research. AODV uses Route Request (RREQ), Route Reply (RREP) control messages in Route Discovery phase and Route Error (RERR) control message in Route Maintenance phase.

In general, the nodes participating in the communication can be classified as source node, an intermediate node or a destination node. With each role, the behaviour of a node actually varies. Mechanism of AODV protocol consists of 3 phases, the route discovery and maintenance of road connectivity management [3]. It must establish and maintain the chosen path is the shortest path with a minimum number of hops as the optimal path without any consideration of the traffic that leads to degradation performance of the network. To improve the quality of service and optimal routing is essential to take into account the distribution of the traffic load in the routing mechanism.

In this work, we have integrated a procedure to avoid congested node. By changing the routing strategy during the route discovery process and also connectivity management. The rest of this paper is organized as follows. Related work is presented in Section 2. Contribution in Section 3, followed by simulation in Section 4. This paper is concluded in Section 5 while the references are given towards the end of paper.

2 RELATED WORKS

The Ad hoc QoS on demand routing (AQOR) protocol has been proposed in [4] that deals with the bandwidth and end to end delay requirements. On demand route discovery, signaling function and hop to hop routing are the main components of the proposed protocol.

QEAO DV [6] routing protocol improves the normal route finding method of AODV for providing QoS in MANETS. QEAO DV establishes a path between the source and the destination on the basis of meeting the application throughput requirement. QEAO DV handles the channel access contention effectively which is the inherent problem in MANET.

The load balancing routing protocols for ad hoc wireless networks can be generally divided into two types based on their basic techniques. The first type is “Traffic-Size” based, in which the load is balanced by attempting to distribute the traffic evenly among the network nodes. The second type is the “Delay” based, in which the load is balanced by attempting to Avoid nodes with high delay. Although our scheme belongs to the “Traffic-Size” based type [9][10].

In LBAR [5], the load metric of a node is the total number of routes flowing through the node and its neighbors. This method is not optimal since it does not account for the various traffic sizes of each route. Associatively Based Routing (ABR) [7] Route is selected based on nodes having associatively states that imply periods of stability. ABR defines a new metric for routing known as the degree of association stability. DLAR [8]: in this protocol is the destination sends the information of the load attached to the RREP packet to the source; after receipt of the packet by the source, The distribution of the traffic load for the AODV protocol [11] is implemented at the route discovery process that modifies the RREQ adds another cost that is based on the number of packets queued all nodes participating in the establishment of the road. The route selection will be based on the minimum cost and the shortest path to destination.

3 CONTRIBUTION

We proposed a several modifications to the AODV:

1. A procedure to verify if the current node is congested or not.
2. Modifying the connectivity management process, when a node sends hello message: it also attaches the value of its load. At the reception of message hello, congestion verification procedure is triggered.
3. Changing the route discovery process.

For the load of a node n_{ij} :

- $L(n_{ij}) = \text{the sum of packets buffered at the queue.}$
Cost in RREQ of the path from the source to destination
- $Cos_load = \text{sum } L(n_{ij}) \text{ of intermediate nodes not overloaded ch to enhance quality of service by}$

distributing the load of traffic in MANETS Novel Approach to enhance quality of service by distributing the load of traffic in MANETS.

The load of node is compared to the cost in RREQ at each reception of message RREQ, that mean If the load of the node is bigger than the cost of active nodes participating at the establishment of road, the current node is considered overloaded.

A node destination receives many paths:

- **Best cost = min (cost_load).**

4 SIMULATION

In this section the performance of the improved version of the protocol is evaluated and compared with the basic version of the protocol. The simulation environment is described and simulation results are presented and discussed.

For each scenario we varied the main parameter that can influence the behaviour and the simulation results is that the pause time for each node representing the immobility time before moving again.

Performance metrics describe variables or simulation input data such as mobility or overload. These metrics are cited: [12]

- **Mobility:** It indicates the movement of nodes. It may be weak or strong. The calculation is done by measuring the relative movement of a node relative to the other.
- **Pause time:** it shows the average time in which the nodes are not moving.
- **Average end-to-end delay:** The average end-to-end Delay is a measure of average time taken to transmit each packet of data from the source to the destination. Network congestion is indicated by higher end-to- end delays.

Table 1. Parameters of simulation.

Time of simulation	900 seconds
Packets size	512 bytes
Rate	4 packets/second
Maximum packet in the queue	70 packets
Number of nodes	50 nodes
Simulator	Ns ₂
Mobility model	waypoint mobility model
Surface	1500 × (300 m
Pause time	0s, 30s, 60s, 120s, 300s, 600s, 900s
MAC	IEEE802.11
Type of file	drop Tail (FIFO)
Application	CBR
Protocol	AODV, AODV_balanced

- Packet Delivery Fraction, in percentage: The fraction of successfully received packets, which survive while finding their destination. Successful packet delivery is calculated such that, all data packets with unique identifier Leaving the source MAC are counted and defined as originating packets. Received packet identifiers are compared to collected transmission data and each unique packet is counted once to ensure prevention of counting excess receptions, which are mainly caused by multiple paths as a result of mobility. The result is the average of the ratio of uniquely received and all uniquely transmitted packets.
- The overhead is calculated by dividing the number of Control packets by received data

packets that criterion illustrates the amount of additive necessary cost for each received data packet.

Simulation scenarios:

- Scenario 1: 20 nodes communicate CBR during 900 seconds, with packets of 512 bytes and a maximum speed of 20 m / s in a field dimensions 1500 m × 300 m.
- Scenario 2: 30 nodes communicate CBR during 900 seconds, with packets of 512 bytes and a maximum speed of 20 m / s in a field dimensions 1500 m × 300 m.

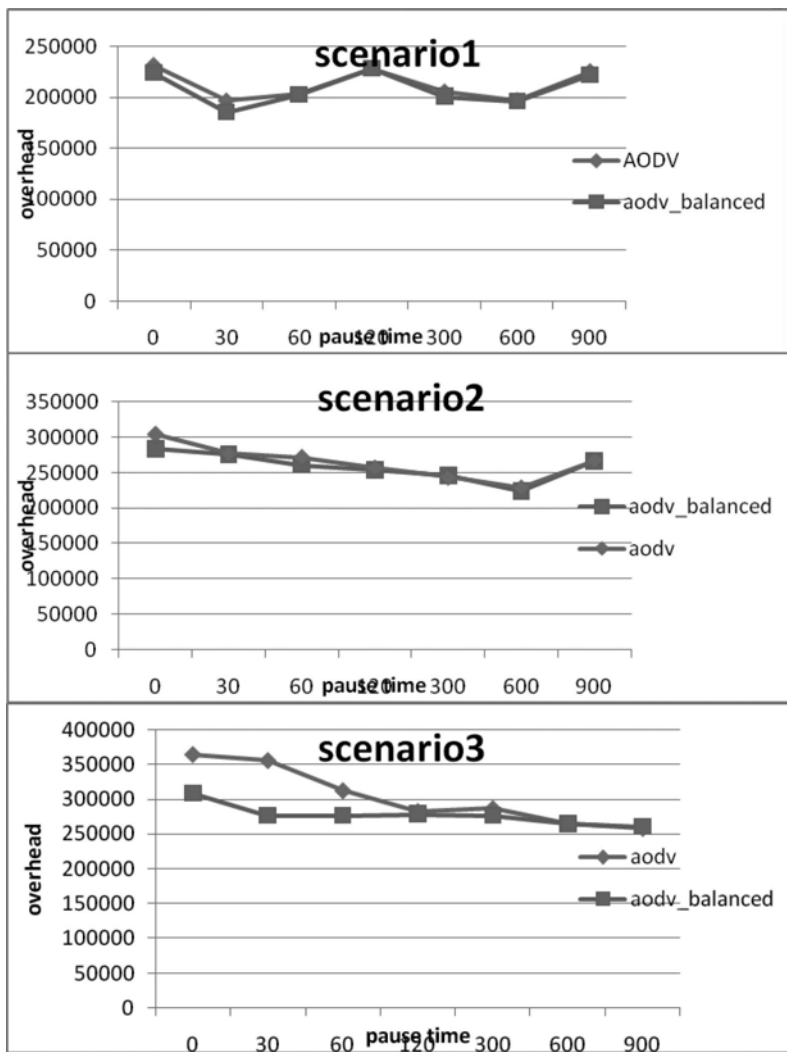


Figure 1. Overhead metric.

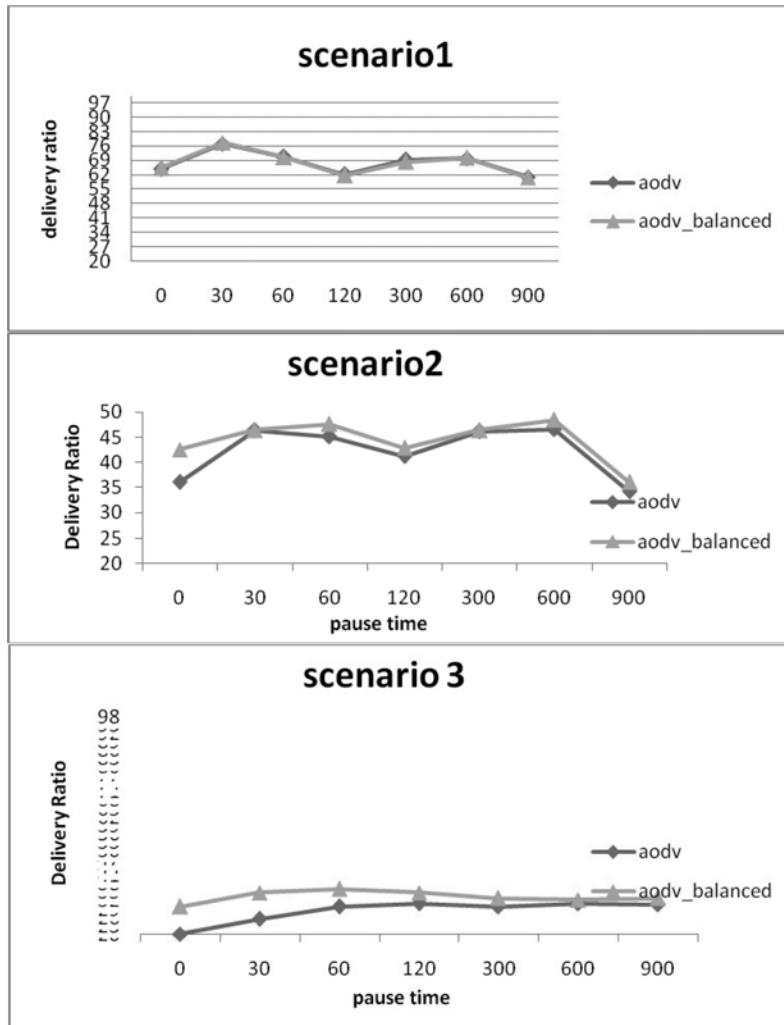


Figure 2 Delivery ratio metric.

- Scenario 3: 40 nodes communicate CBR during 900 seconds, with packets of 512 bytes and a maximum speed of 20 m/s in a field dimensions 1500 m × 300 m.

Discussion of results:

The results of each scenario for this metric based on Pause-Time is shown in the following graph:

1. Overhead:

In this Figure 1: There has been a decrease in the three scenarios on average 3% is due to the increase in the numbers of received packets. Which is due to the mechanism of distribution of traffic load? And the Figure also shows that the overhead of two

protocols increases with increasing the number of sources. This is because the increase in the number of source nodes causes a greater number of request messages flooding. In contrast, AODV_balanced adopts a mechanism for load balancing, which tries to route packets along a less congested path to avoid overloading some nodes. In terms of overload AODV_balanced is more efficient than AODV.

2. Delivery ratio:

In scenarios 1, 2 and 3, AODV_balanced achieves the highest packet delivery fraction for all pause timevalues. This factor is influenced by the removal of packets at the queue, if it reached the maximum number. In our simulation, the maximum number is 70 packets. The modified protocol is more efficient than AODV because the

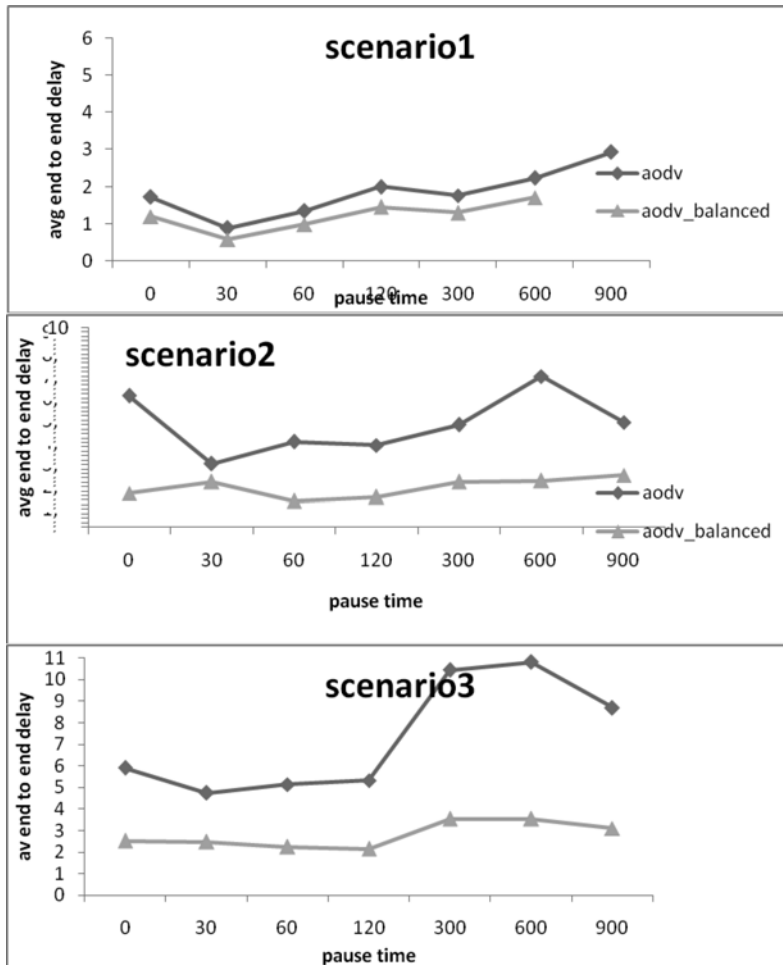


Figure 3. Average end-to-end delay.

verification process of the size of the queue. The value of the size is compared with the cost of the request at each receiving a request, if it is higher, the node is considered overloaded, the request RREQ will be deleted, the queue will never reach the maximum size so fewer lost packets, and the rate of delivery will be better for modified AODV.

3. Average end-to-end delay:

This figure shows:

In the case of the simulation of scenario 1, the average delay of Aodv_balanced is an average of 2% lower than AODV. In the case of the simulation of scenarios 2, 3 the average delay of aodv_balanced is an average of 3 and 5% lower than AODV.

AODV_balanced achieves significantly lower delay than AODV. The modified AODV is more efficient than AODV. It carries a lower delay than

the AODV in scenarios 2 and 3 we have less load on the nodes in the network and the waiting time at each node has decreased. Aodv_balanced increases the network throughput.

5 CONCLUSION

In the paper, Aodv_balanced is presented based on the original AODV, The route discovery mechanism and connectivity management of basic AODV are modified, New route discovery mechanism selects routes for transmission. The simulation results show that Aodv_balanced has a better end to end delay and a higher rate of delivery than AODV and can improve the QoS in large Mobile ad hoc networks.

REFERENCES

- [1] www.ietf.org/html.charters/manet-charter.html.
- [2] T. Bheemarjuna Reddy, I. Karthigeyan, B.S. Manoj, and C. Siva Ram Murthy. Quality of service provisioning in ad hoc wireless networks: a survey of issues and solutions. Elsevier B. V., volume 4 (ISSN: 1570-8705/05/\$).
- [3] C. Perkins and E. Royer, "AD hoc on-demand distance vector routing".
- [4] Qi Xue and Aura Ganz "Ad hoc QoS on-demand routing (AQOR) in mobile ad hoc networks" *Journal of Parallel and Distributed Computing*,(63), 154–165, 2003.
- [5] V. Saigal, A. K. Nayak, S. K. Pradhan, and R. Mall, "Load Balanced routing in mobile ad hoc networks", *Elsevier Computer Communications* 27(2004), pp. 295–305.
- [6] D.S. Thenmozhi and R. Lakshmi pathi "Quality of Service Enhanced Routing in Mobile Ad Hoc Network" *International Journal of Computer Science and Information Security*, Vol 8, No. 5, August 2010.
- [7] Chai-Keong Toh "Associativity-Based Routing for Ad-Hoc Mobile Networks" *Wireless Personal Communications* 4: 103–139, 1997. 1997 Kluwer Academic Publishers. Printed in the Netherlands.
- [8] Sung-lu and Mario, "dynamic load aware routing" helsinki. *IEEE ICC Finland*.
- [9] S. Bharadwaj, V. Kumar, A. Verma 3 the "review of load balanced routing protocol". *International Journal of Engineering Trends and Technology- July to Aug Issue* 2011.
- [10] Lee, Y.J. and Riley, G.F., "A Workload-Based Adaptive Load-Balancing Technique for Mobile Ad Hoc Networks," *IEEE*, pp. 2002–2007 (2005).
- [11] A. Guidoum, A. Boukelif, Optimization of AODV routing protocol in mobile ad-hoc network by introducing features of the protocol LBAR, *Proceedings of the 5th European Conference of Communications (ECCOM '14) Geneva, Switzerland, December 29–31, 2014*.
- [12] C.E. Perkins, E.E. Royer, S.R. Das, and M.K. Marina, "Performance comparison of two on demand Routing protocols for ad hoc networks," *In IEEE Personal Communications*, Feb 2001, vol.

Artificial intelligence in e-learning

Hachem Alaoui Harouni, Elkaber Hachem & Cherif Ziti

*Research Team EDP and Scientific Computing, Mathematics and Computer Department,
Faculty of Science, Moulay Ismail University, Meknes, Morocco*

ABSTRACT: This paper aims at suggesting solutions to some problems that E-learners face through investigating the realization of an e-learning multi-agent system which is a part of an artificial intelligence field that treats the acoustic parameters deduced from the voice of the e-learner. We can determine the emotional status of the e-learners by analyzing the parameters of the voice. The results of this analysis are fundamental to help us solve many problems that most of e-learners face, such as: isolation and boredom. Basing on the notion of agents we can realize a system named ASTEMOI, that is consists of three agents, and based on a client-server architecture that utilizes the well-known method, SVM, which can crack the problems of classification so as to determine the suitable and the unsuitable states of the study, and we use the Logistic Regression method to estimate the strengths and weaknesses of a learner in a certain class with the help of the information collected from ancient undergraduate students from the faculty of Chariaa of Fez.

We are going to describe our system and how we can derive the emotions from the feedback and the analysis of the voice.

1 INTRODUCTION

To cope with the requirements of ICT, faculties must now adapt developed systems of learning and use them to know the potential of their students and cultivate them in order to become major assets, not only to obtain a diploma, but also to boost their set of skills and competences and advance in their fields.

The teachers' concern is expressed by a desire to help or offer a service or a training of good quality. Teachers must put the student at the center of their concerns and identify their needs and expectations. This skill requires a commitment to add value to any action in order to meet and exceed the students' expectations. In order to make this happen, we can add a diagnosis table of the courses being taught.

The role of such a table is to give us estimation about the difficulties and hardships that hinder the learning process for students and, therefore, help the teacher enhance the quality of the courses and make them more suitable for them.

2 EMOTION, AGENT, SYSTEM ASTEMOI

2.1 Emotion

Emotion is a quick process focused on an event and consists of a trigger mechanism based on

the relevance that shapes a multiple emotional response (Pasquier & Paulmaz. 2004). The emotion was long regarded as opposed to cognition. So many philosophers including Plato, Descartes and Kant, consider this phenomenon as a disturbance of reason that it was absolutely necessary to correct. For them, rationality and reason should not give way to emotions. In this line of thinking, most theories of education focused on the development of cognitive processes and neglecting the emotional dimension (Talhi S. 1996), yet emotions color the events of life, give them value and are an engine, a motivation (Baron M. Février 2001). Furthermore, emotions can interfere and make a difference between all the elements. And that's the difference between a "cold", rational, and a "hot" cognition, emotional (Behaz A., Djoudi M. & Zidani A. 2003).

2.2 Multi-agent system

A Multi-Agent System (M.A.S.) is a computerized system composed of multiple interacting intelligent agents within a certain environment. Multi-agent systems can be used to solve problems that are difficult or impossible for an individual agent or a monolithic system to solve. Intelligence may include some methodic, functional, procedural approaches, algorithmic search or reinforcement learning (Niazi Muaz & Hussain Amir 2011).

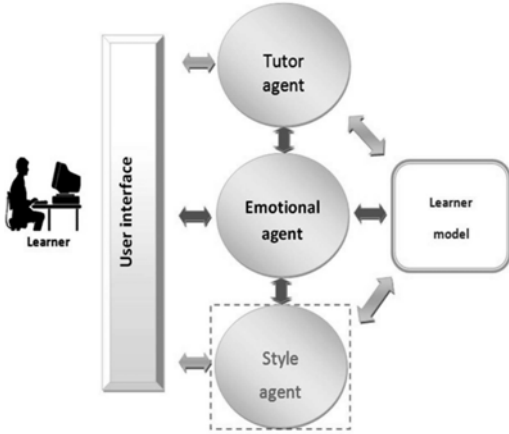


Figure 1. Global system architecture.

2.3 The system ASTEMOI

To guarantee that the quality of distance learning is being used to present the concept of agents in an Emotionally intelligent Tutorial System called ASTEMOI whose architecture is as follows:

- **Tutor agent:** Is the agent in charge of managing the courses and cognitive status of the e-learner. (Brusilovsky P. 2001)
- **Style agent:** Using the Felder Questionnaire (Derouich Aziz. 2011) we can determine the suitable learning style of the e-learner.
- **Emotional agent:** The role of this agent is to determine the emotional status of the e-learner using the results of the voice analysis and the feedback.

3 THE PROBLEMS THAT THE ASTEMOI SYSTEM TREATS

The great majority of digital natives face many problems when it comes to learning using E-learning applications or websites. One of the problems that they face is the fact that these lessons are not tailored to suit neither their cognitive abilities nor their learning style. Another problem that they encounter is blockade and isolation. The online learning resources do not provide their customers with the motivation and interaction they need in order to perform well and they do not take their emotions into account.

Since every learning situation's central objective is to provide the learners with good-quality training, the ASTEMOI system highly considers the learners' needs and also their strengths and weaknesses, because they are the main factors to boosting e-learners performance and improving the quality of e-learning applications.

4 ACOUSTIC PARAMETERS CHARACTERIZING THE EMOTION

The results of several studies show that the changes in speakers' state modify specifically the acoustic parameters of their word (Picard R. 1997).

Acoustic analysis of vocal emotion essentially depends on the following parameters: fundamental frequency (F_0), intensity and duration of the emotional voice. The values obtained directly reflect the physiological changes of the speaker, who feels particular emotions in particular situation (Skinner 1935). To synthesize emotional speech, the descriptors of the quality of voice called high-level (intensity and frequency) are the most used as they provide a high level of interpretation (P-Y. Oudeyer 2003).

- **Fundamental frequency F_0 (or pitch):** a particularly relevant index in the expression and perception of emotion. It is related to the pitch of the voice (acute or severe). The signal is modeled as the sum of a periodic signal T and a white noise, such as:

$$F_0 = \frac{1}{T} \quad (1)$$

To estimate the fundamental frequency, several methods are available; one chooses the method used by the Praat software (Boersma P & D. Weenink. 2005) and that consists in searching for similarities between the shifted versions of the observed signal, denoted s , defined as follows.

$$r_s(m) = \begin{cases} \frac{\sum_{n=0}^{N-1-m} s(n)s(n+m)}{\sqrt{\sum_{n=0}^{N-1-m} [s(n)]^2} \sqrt{\sum_{n=0}^{N-1-m} [s(n+m)]^2}} & \text{if } m \geq 0 \\ r_s(-m) & \text{if not} \end{cases} \quad (2)$$

The period T is estimated by finding the smallest value of m for which $r_s(m)$ is maximum.

- **The intensity I :** provides a measure of the loudness of the voice (low or high), it is calculated on a signal portion of length N having the following form:

$$I = 10 \log \left[\sum_{n=1}^N w(n)[s(n)]^2 \right] \quad (3)$$

where w is the Gaussian analysis window (Amir N & S. Ron 1996).

- **The flow of speech (Q noted):** one of the parameters calculated with F_0 in the description of the

Table 1. The acoustic indices of the various primary emotions.

	Frequency domain	Temporal domain	Voice quality
Joy	<ul style="list-style-type: none"> - F₀ medium high - Variation of F₀ dynamic 	<ul style="list-style-type: none"> - Fast flow - rhythmic structure regular accentuation 	<ul style="list-style-type: none"> - Depth high but not as much as for anger - Voice Slightly sucked
Anger	<ul style="list-style-type: none"> - F₀ high or moderate means - Change in the dynamic F₀ - Outline F₀ down heavily at the end of sentence 	<ul style="list-style-type: none"> - Flow faster than the neutral voice but less than for joy 	<ul style="list-style-type: none"> - High depth - Voice drawn and taut - Great Energy in high frequencies
Sadness	<ul style="list-style-type: none"> - Average F₀ at the neutral voice - Little change in F₀ 	<ul style="list-style-type: none"> - Slow flow - Rhythm with regular breaks 	<ul style="list-style-type: none"> - Low depth without variation - Articulation Less accurate
Fear	<ul style="list-style-type: none"> - F₀ slightly medium high - Large variation of F₀, but not as much as joy or anger 	<ul style="list-style-type: none"> - slower flow than for the joy and anger but faster than the neutral voice - Irregular pauses 	<ul style="list-style-type: none"> - low intensity - Accurate articulation - Low energy at low frequencies

vocal emotion which is the number of syllables per second.

$$Q = (NSE \times 1000) / \Delta t \tag{4}$$

With:

NSE: The number of syllables in the statement.
 Δt : The duration of the statement.

The variation of the acoustic parameters of the emotional voice is often described in terms of the degree of deflection of their values that are relative to the values found in the neutral voice. Some acoustic characteristics of emotions, considered primary, are shown in the following table: (Chung Soo-Jin. 2000)

We are going to limit our study in only two states; favourable (Joy...), and unfavourable (Fear, Sadness, Anger ..) in order to detect the emotional states that are positive for the e-learning process, using the well-known method, SVM.

5 AUTOMATIC RECOGNITION OF EMOTIONS

Speaking is a medium that contains not only linguistic information but also provides information on personality traits and the emotional state of the speaker.... Such information can be exploited

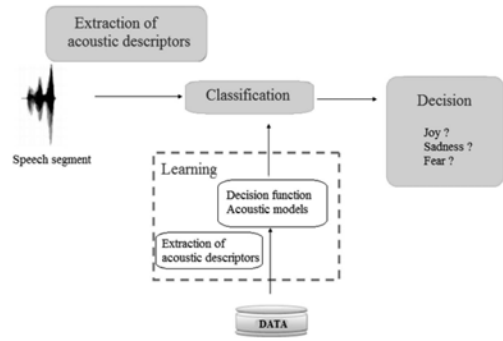


Figure 2. Principle of an emotion recognition system.

technically to allow the machine to understand human speech. In recent years studies on the emotional speech begin to address the development of automatic classification system of emotions, based on four main phases (Chloé Clavel 2007):

- a. Extraction of acoustic descriptors: This step has a role to transform the speech signal into a sequence of acoustic vectors containing the various descriptors used to make a representation of the main acoustic characteristics of the speech signal.
- b. During learning: several acoustic vectors corresponding to the same sounds in the same class can be grouped to a representative of this class.
- c. During the classification phase a comparison between the acoustic vectors of the speech signal to be analyzed and representatives of each class or models leads to achieve a probability of belonging to each class for each acoustic vector.
- d. The decision phase is the exploitation phase of probabilities calculated to associate a class to a speech segment.

6 MACHINES LEARNING (SUPPORT VECTOR MACHINE)

There are many methods that can resolve the problem of classification. In our case, the most suitable method is the SVM, inspired by the statistical learning theory. The objective of this method is to segment the collected data into two divisions by maximizing the distance between them. To generalize, afterwards, multi-classes (C.W. Hsu & C.-J. Lin).

In our case, this method is going to divide the learners into two states: one is when the learning process is effective and is taking place and the other one is when it doesn't.

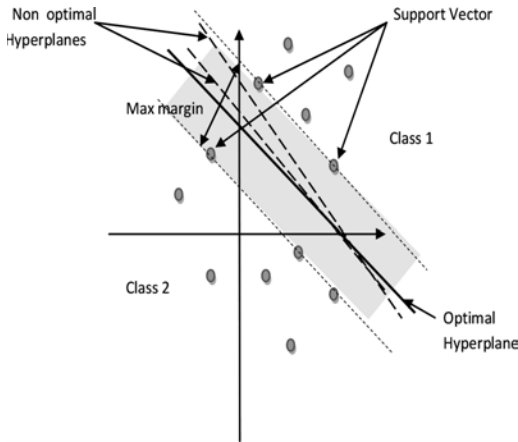


Figure 3. Hyperplane separating data belonging to two classes.

7 THE PREDICTIVE MODEL

It is very necessary to determine the learners' most suitable learning style, that's why it is highly recommended to give them sufficient time to answer the questionnaire of Felder (R.M. 1993). They even can be provided with a personal profile that contains their strengths and weaknesses in a given discipline or course. This style, ASTEMOI system, has an analytical representation that aims at achieving that goal. That's why we must take the following steps:

- Using the data collected from ancient students to shape the central part of the model relying on the logistic regression Beta-Bernoulli method.
- Relying on the data collected from alumni students to estimate the new students' strengths and weaknesses.
- Redirecting students to the most suitable profile that would respond to their motivation and grant them with extra links to courses suggested by the tutor agent. The courses can take the forms of videos, audios or a text according to the student's learning style (Reflection, Reasoning, Sensory, and Progression).
- Comparing the final results of learners with the results estimated by the model which allows for improving and updating the predictive model.

8 LOGISTIC REGRESSION

We use this process to estimate whether the e-learner faces any difficulty or not in a certain class. The method is called the Logistic

Regression Beta Bernoulli inspired by the mathematical model of the Bernoulli distribution; it adjusts to the experiments in which the results could have two values 0 or 1. In our work we will identify factors related to having problems in a subject for a student characterized by the following attributes:

- Type Bac: Arts, Sciences, Islamic studies.
- Age: [19,24]; [24,30]; > 30
- Mark: <10; [10,12]; >12

The mathematical background is similar to the model in which Y is a binary variable (0 for non-occurrence of the event; and 1 when the event occurs) with Y random and X_i not-random.

Let Y variable to predict (explained variable), $X = (X_1, X_2, \dots, X_i)$ the predictor variables (explanatory variables). So that the expectation of Y takes only two values, we use the logistic function: (Boudin F. 2012)

$$f(x) = \frac{\exp(x)}{1 + \exp(x)} = p \quad (5)$$

So $0 < f(x) < 1$ and $E(Y) = 0$ or 1 .

We distinguish two cases:

- **The first** case is that of a single variable, the values of x and y describe each possibility namely $X = 0$: No criterion; $X = 1$: existing criterion; $Y = 0$: having problem (no) $Y = 1$: having problem (yes). So we have: (Neji Sonia & Jigorel Anne-Hélène 2012)

$$\text{Logit}[PY_i = 1 | X = x] = \beta_0 + \beta_1 x \quad (6)$$

- **The second** case is the multiple logistic model. The variables used to establish the link between multiple cases.

$$\begin{aligned} \text{Logit } P(\text{Problem} = \text{yes} | \text{Age}, \text{Mark}, \text{Type_Bac}) \\ = \beta_0 + \beta_1 \cdot \text{Age} + \beta_2 \cdot \text{Mark} + \beta_3 \cdot \text{Type_Bac} \end{aligned} \quad (7)$$

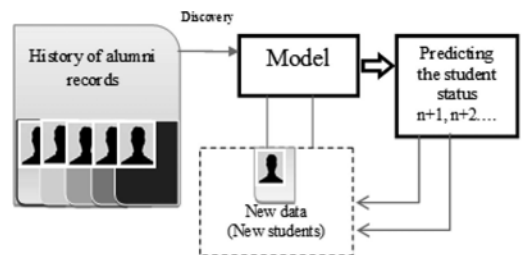


Figure 4. Predictive model in the system ASTEMOI.

So we led to the probability to estimate students who are going to have problems in a subject or not.

9 CONCLUSION

Thanks to e-learning multi-agent system, we are able to detect the learners' emotions using automatic systems during a distant training. These systems enable us to determine the emotional states of E-learners. This paper opens up the doors to future researches that may adopt other systems and methods to provide more feasible results, like treating and analyzing facial expressions using detectors, or even adding other models that would help detect the psychological state of an e-learner through his/her interaction with the online course, which is going to help us find the right motivation for each learner and then ameliorate the result. They can also ameliorate the predictive system, ASTEMOI, by adding more devices that will help orient the learners in their academic careers and even in choosing the topic of their research projects.

REFERENCES

- Amir N. & S. Ron (1996). Towards an automatic classification of emotions in speech. Dans Proc. of ICSLP, Philadelphia.
- Baron M. Février (2001) "Intelligence Artificielle EIAH", Ecole et Science Cognitives.
- Behaz A., Djoudi M. & Zidani A. (2003). « Approche de modélisation et d'adaptation des documents pédagogiques hypermédias en enseignement à distance », Actes du 6ième colloque CIDE'6, Caen, France.
- Boudin F. (2012). Machine Learning avec WEKA module X8II090 course 1, Department of computer science, Nantes university.
- Boersma P. & D. Weenink. (2005). Praat "Doing phonetics by computer [computer program], from <http://www.praat.org/>." Rapport Technique.
- Brusilovsky P. (2001). "Adaptive hypermedia, in user modeling and user adapted interaction".
- Chung Soo-Jin. (2000). "L'expression et la perception de l'émotion extraite de la parole spontanée" évidences du coréen et de l'anglais.
- Chloé Clavel. (2007). "Analyse et reconnaissance des manifestations acoustiques des émotions de type peur en situations anormales".
- Derouich Aziz. (2011). Conception et réalisation d'un hypermédia adaptatif dédié à l'enseignement à distance.
- Hsu C.W. & C.J. Lin. A comparison of methods for multi-classe support vector machines. IEEE Transactions on Neural.
- Neji Sonia & Jigorel Anne-Hélène. (2012). La regression logistique, exposé statistiques et économétrie.
- Niazi Muaz & Hussain Amir. (2011). "Agent-based Computing from Multi-agent Systems to Agent-Based Models: A Visual Survey" (PDF). *Scientometrics* (Springer) 89 (2): 479–499. doi:10.1007/s11192-011-0468-9.
- Oudeyer P.Y. (2003). The production and recognition of emotions in speech: features and algorithms. *International Journal of Human Computer Interaction*, special issue on Affective Computing, 59 (1–2): 157–183.
- Pasquier & Paulmaz. (2004). « La gestion des émotions et les implications dans l'apprentissage ».
- Picard R. (1997). *Affective Computing*. MIT Press, Cambridge, MA.
- R.M. (1993). Reaching the Second Tier: Learning and Teaching Styles in College Science Education, *J. College Science Teaching*, 23(5), pp. 286–290.
- Skinner (1935); Fairbanks & Hoaglin, 1939, 1941; Black, 1961; Williams & Stevens, 1972; Cosmides, 1983; Laukkanen et al., 1996; Leinonen et al., 1997.
- Talhi S. 2–4 avril, (1996). "Moalim: un système auteur de l'EIAO", Actes du 18ième symposium DECUS France, Paris.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Highlighting the evaluation gaits in the adaptive learning system ALS_CORR[LP]

N. El Ghouch & E.M. En-Naimi

LIST Laboratory, Faculty of Science and Technology of Tangier, UAE, Tangier, Morocco

Y.Z. Seghroucheni, B.E. El Mohajir & M. Al Achhab

The Faculty of Sciences, UAE, Tetuan, Morocco

The National School of Applied Sciences, UAE, Tetuan, Morocco

ABSTRACT: The aim of this paper is to spot the crucial moments of evaluation in the adaptive learning system ALS_CORR[LP] (Adaptive Learning System_CORRection[Learning Path]). In general, the adaptive learning systems are well known for providing a specific learning path according to the prerequisites, learning styles, Etc. But their main disadvantage remains to treat the case, where a generated learning path is not the leading one. Therefore, and since the evaluation is the only way to consider the success of a learning, knowing the moments where the evaluation is required, allows optimizing the system performance, by taking into consideration the possibility of correcting the non-leading paths which is the case of the studied system through this paper.

1 INTRODUCTION

The use of Internet for educational purposes is currently growing in many forms. This fact has enabled the emergence of several e-learning systems. The adaptive learning systems are one class of those systems that knows a major success, this achievement is due primarily to the endless possibilities they offer in terms of personalizing learning paths according to the prerequisites, objectives, learning styles, etc However those systems consider work is done, once the learning objects are assigned, and do not offer any corrections neither for the generated learning paths or the learners profile, since the learning objects are linked essentially to them. The assessments remain the only way to achieve any eventual correction, that's why we put the focus in this paper to the essential moments where the evaluation must intervene inside system, and how it could be helpful for a reconsideration of the learner profile.

The rest of this paper is organized as follows:

In section 2 we will be discussing the evaluation and highlighting its different types, then we will put the focus on the adaptive learning systems and evoke the characteristics of the system studied in this paper, later in section 4 we will uncover the crucial moments of evaluation in an adaptive learning system, finally some conclusions are drawn in the last section.

2 THE EVALUATION

Evaluation can have many purposes, including collecting feedback for performance improvement, or Gathering data, in order to take a decision, because The more information we have about learners, the clearer the picture we have about achievement or where gaps may occur.

Evaluation can be characterized as being either formative or summative (see Figure 1).

Broadly, formative evaluation looks at what leads to an intervention working (the process), whereas summative evaluation looks at the short-term to long-term outcomes of an intervention on the target group. Evaluations are normally divided into two broad categories: formative and summative.

The terms formative and summative do not have to be difficult, yet the definitions have become confusing in the past few years. This is especially



Figure 1. The types of evaluation.

true for formative assessment. In a balanced assessment system, both summative and formative assessments are an integral part of information gathering.

2.1 Summative evaluation

Summative Assessments are given periodically to determine at a particular point in time what students know and do not know. Many associate summative assessments only with standardized tests such as state assessments, but they are also used at and are an important part of district and classroom programs. Summative assessment at the district and classroom level is an accountability measure that is generally used as part of the grading process. The list is long, but here are some examples of summative assessments:

- State assessments
- District benchmark or interim assessments
- End-of-unit or chapter tests
- End-of-term or semester exams

2.2 Formative evaluation

Formative Assessment is part of the instructional process. When incorporated into classroom practice, it provides the information needed to adjust teaching and learning while they are happening. In this sense, formative assessment informs both teachers and students about student understanding at a point when timely adjustments can be made. These adjustments help to ensure students achieve targeted standards based learning goals within a set time frame. Although formative assessment strategies appear in a variety of formats, there are some distinct ways to distinguish them from summative assessments.

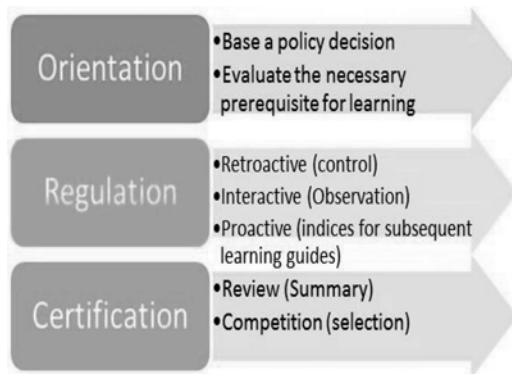


Figure 2. The objective of evaluation.

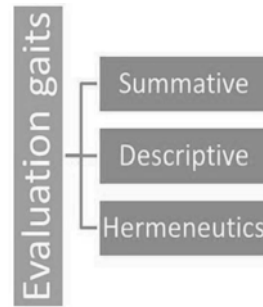


Figure 3. The different approaches of evaluation.

2.3 The evaluation, Why?

The generic goal of most evaluations is to provide “useful feedback” to a variety of audiences including most importantly learners, tutors and the system administrator. Therefore the evaluation has three major goals: Orientation, Regulation and Certification which are explained in the figure below:

The evaluation happens mostly according one of the following approaches

- Summative: Measurement
- Descriptive: Description of behavior, performance, challenges, products, procedures
- Hermeneutics: Intuitive and subjective interpretation of a body of evidence (to direct, control, certification)

3 THE ADAPTIVE LEARNING SYSTEM

An Adaptive Learning System (ALS) is a Learning Management System (LMS) that, quite simply, can adapt to the needs of the learner. Unlike a traditional LMS, which acts as a repository of information and a tool for the training administrator to assign modules and track progress, an ALS assigns modules based on learner needs/styles/competence level/etc. And sometimes it has the ability to assess learner progress and account for this while creating the learning path. There are several approaches that fall into the direction of personalizing learning path and offering an adapted content to the learner profiles, those works can be summarized into two categories:

The first category contains systems that tend to use implicit methods for identifying learning styles based mainly on the analysis (Carchiolo, Longheu, Malgeri, & Mangioni 2007), (Bousbia, Rebaï, Labat, & Balla 2010) and observation (Bousbia, Rebaï, Labat, & Balla 2010), (Graf & Kinshuk 2007) of the learners behaviors in the system,

however those methods are not completely reliable given the fact that the learners can engage in other activities during learning.

The second category contains the content adaptation systems that use explicit methods for identifying learning styles by using e-questionnaires (Tzouveli, Mylonas, & Kollias 2008), (Bontchev & Vassileva 2012), (Graf, Viola, Leo, & Kinshuk 2007) or letting the learners express their preferences (Guerrero-Roldán & Alfonso 2007) personal characteristics (Jean-Daubias) or using the Felder-Silverman learning style model (Felder & Silverman 1988), (Papanikolaou, Grigoriadou, Kornilakis, & Magoulas 2003).

The problem with those two types of systems cited previously, is that once the learning path is generated, it is supposed to be as automatically the leading one, which is not always the case, since we can always notice failure in the assessments.

The adaptive learning system studied in this paper is the one cited on the works of (Seghroucheni, Mohajir, et al. 2014). This system has the ability to adapt itself to a situation of a learners failure in any assessment, by offering the possibility to correct the learner profile or even in other times to recommend the most relevant learning objects.

Here is an overview of the ALS_CORR[LP] system.



Figure 4. Overview of the homepage (1/2).



Figure 5. Overview of the homepage (2/2).

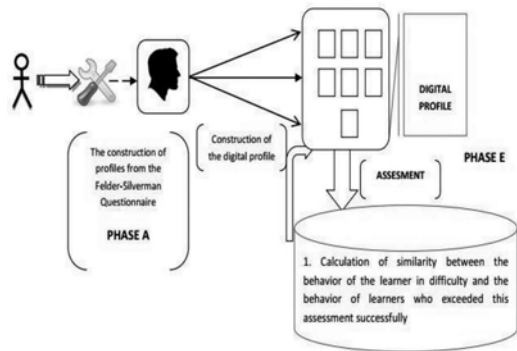


Figure 6. The learning scenario in the adaptive learning system.

In facts the learning process takes place according to the following scenario:

For a first-timer, the learner must respond to a prerequisites test and fill the questionnaire of Felder-Silverman learning style model known as FLSLM, in order to determine his initial theoretical profile.

The next step is to allocate an appropriate version of the actual course according to the compatibility with his learning style and the test result.

The system will correct the learning path of the learners who have obtained non qualifying score by recommending the learning path of those who have passed successfully the assessment and have the same initial profile. This recommendation will be based on the calculation of similarity between the behavior of the learner who has failed at the assessment and the behavior of other learners on the system. If the similarity does exist, the learner in difficulty will be proposed for the current course the same versions of learning object and consequently the same learning path of the one with whom he has a similarity in the behavior.

3.1 The learning styles of Felder-Silverman

In (Nunan & Lamb 1996) the authors have presented the four dimension of Felder-Silverman Learning Style Model (FSLSM), each learner is characterized by a particular preference for each of these dimensions:

Active / Reflective: How do you process information?

Active: They learn by doing something with the information. They prefer to process information by talking and trying the subject of learning.

Reflective: They think about the information. They mostly prefer to understand before acting.

Sensing / Intuitive: How do you take the information?

Sensing: They prefer to take information that is concrete and practical. They have a sense of detail, facts and figures and prefer to use proven procedures. They are realistic and like practical applications.

Intuitive: These learners prefer to take information that is abstract, original, and oriented theory. They look at the big picture and try to understand the general trends. They like to discover possibilities and relationships between ideas.

Visual / verbal: How do you prefer information to be presented?

Visual: Visual learners prefer visual presentations of material diagrams, chart, graphs, and pictures.

Verbal: Verbal learners prefer explanations with words- both written and spoken.

Sequential / Global: How do you prefer to organize information?

Sequential: Sequential learners prefer to organize information in a linear, orderly fashion. They learn in logically sequenced steps and work with information in an organized and systematic way.

Global: Global learners prefer to organize information more holistically and in a seemingly random manner without seeing connections. They often appear scattered and disorganized in their thinking yet often arrive at a creative or correct end product.

3.2 *The index of learning style*

The Index of Learning Styles (ILS) developed by Felder and Soloman, is a questionnaire of 44 items to identify learning styles according to Felder-Silverman. As mentioned earlier, each student has a personal preference for each dimension. These preferences are expressed with values ranging from +11 to -11 per dimension, with steps + / -2. This range has eleven questions that are asked for each dimension. In response to a question, for example, with an active preference, one is added to the value of the active/Reflective dimension while a response to a preference Reflective decreases the value of 1.

Therefore, each question is answered either with a value of 1 (answer a) or -1 (answer b). Answer a is a preference for the first pole of each dimension (active, sensing, visual, or sequential), answer b is to the second pole of each dimension (Reflective, Intuitive, verbal or Global). The ILS is an index often used and well studied to identify learning styles. In (Felder & Spurlin 2005) the authors gave an overview of studies on the analysis of data from

the ILS as regards the distribution of preferences for each dimension and to check the reliability and validity of the index.

3.3 *Analyzing the outcomes of the scenario*

In the figure above, there are three major phases:

- Phase A: the phase where the Felder-Silverman profile is constructed.
- Phase B: the observation of the learners behavior.
- Phase E: this phase represents the assessment.

According to the Figure 6, the possible scenarios are:

1. If $(A \equiv B) \Rightarrow E$ (if the initial profile in the platform, match the alleged behavior to be adopted by the learner, and the result of the assessment is positive).
2. If $(A \equiv B) \Rightarrow \neg E$ (if the initial profile in the platform, match the alleged behavior to be adopted by the learner, and yet the result of the assessment is negative).
3. If $(A \neq B) \Rightarrow E$ (if the initial profile in the platform doesn't match the alleged behavior to be adopted by the learner, and the result of the assessment is positive)
4. If $(A \neq B) \Rightarrow \neg E$ (if the initial profile in the platform doesn't match the alleged behavior to be adopted by the learner, and yet the result of the assessment is negative).

The cases of interest are: if $(A \equiv B) \Rightarrow \neg E$ and if $(A \neq B) \Rightarrow \neg E$ because the recommendation is only to learners experiencing difficulties in learning.

The $(A \equiv B) \Rightarrow \neg E$ case The proposed solution is to calculate the similarity between the behavior of the learner in difficulty with the behavior of other learners who have the same theoretical profile and having successfully exceeded the assessment in question, and recommend subsequent path of learning to him; this similarity is based on the items described in the Table1.

The $(A \neq B) \Rightarrow \neg E$ case This specific case shows that there is clearly a problem with the course itself, and its up to the tutor himself to reevaluate the stages of the course and its didactic transposition verret.



Figure 7. The different aspects of evaluation.

4 THE EVALUATION MOMENTS IN THE ADAPTIVE LEARNING SYSTEM ALS_CORR[LP]

Judging from the learning scenario detailed in the last section, we can conclude that there are two types of evaluation happening in an ALS that aims to correct the learning path: an implicit evaluation, which can be assigned to a formative evaluation, and an explicit evaluation (summative).

Lets start first by spotting when the evaluation is taking place; in fact there are three main moments where the evaluation seems to be crucial and inevitable, all in all to ensure the best system performance. Those moments are highlighted in the figure below:

4.1 Moment 1

The first evaluation moment is the prerequisites test, which is, together with the Felder-Silverman test, builds a solid starting point of the learning process. Here is in the following figures a screenshot of the two evaluations moment in the system.

As a matter of fact, the parameters listed previously, serve to build a recommendation system that will operate according to the following steps [12]:

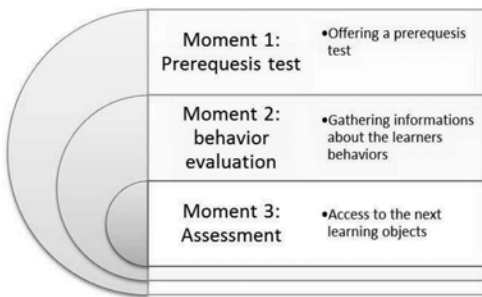


Figure 8. The evaluation moments in an ALS.



Figure 9. The Felder-Silverman questionnaire.



Figure 10. The prerequisites test.

Table 1. The list of the parameters referring to the behaviors.

Designation	Signification
NBREXR	The performed exercises
NBREXM	The studied examples
NBRTST	The made assessment
ORDRPA	The LO traversal order
TMPTH	The theoretical part duration
FC	The connection frequency
TCE	the connection timing vs Assessment
DP	The participation degree in chat.
TS	each session duration



Figure 11. The system dashboard.

4.2 Moment 2

The second evaluation moment is where the system starts gathering informations about the conduct of learning, as a matter of fact, the authors of [12] listed the parameters referring to the learners behavior, those parameters can be listed in the following table:

The objective of this evaluation is to provide enough information to correct either the learning path or the learner profile in the case of a failure in an assessment. Here is a screenshot of the page allowing monitoring the system and gives access to the learners behavior.

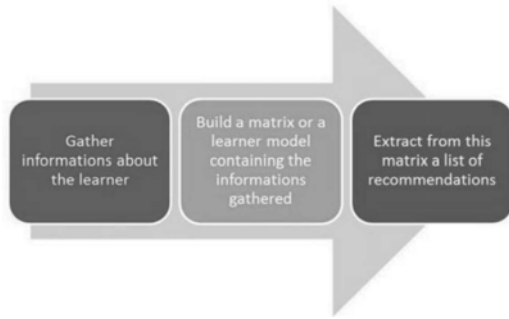


Figure 12. The steps of the recommender system.

Gather informations about the learner

A distinction can be made between two forms of data collection:

Explicit data collection—Active filtering: based on the fact that the learner explicitly tells the system his interests in learning preferences, media, etc...

Implicit data collection—passive filtering: based on observation and analysis of the learner behavior, made implicitly in the application that embeds the recommendation system, everything is done in background. This is the level where we the studied system is operating, by first identifying learners with the same behavior as that of the learner experiencing difficulties.

Learner model

The user model is generally in the form of a matrix. It can be represented as a table that contains data about the learner behavior, which varies by the way according to the items listed in table 1.

List of Recommendation

To retrieve a list of suggestions from a user model, the algorithms use the concept of similarity measure between objects or persons described by the model learner. The similarity aims to provide a value or a number (in the mathematical sense) to the similarity between two things. The stronger the similarity is, the bigger the value of the similarity will be. Conversely, the weaker the similarity is, the smaller the value of the similarity will be. The conventional approach for recommendation systems is to build models of users based on information about them. In this system we are talking about the elements of Table 1.

For 2 learners (U_i) and (U_j):

$$\begin{aligned} \text{Pred}(U_i, U_j) = & \alpha1(TS) + \alpha2(FC) + \alpha3(DP) \\ & + \alpha4(TMP TH) + \alpha5(TCE) + \alpha6(NBREX) \\ & + \alpha7(NBREXM) + \alpha8(NBRTST) \\ & + \alpha9(ORDRPA) \end{aligned}$$

Each learner can be considered as an incomplete vector which we know only a few components. However, it is possible to calculate a similarity between such vectors by restricting to only components they have in common.

Assuming that the behavior of learners U_i and U_j are random variables X_i and X_j after an unknown joint distribution, it is possible to define the correlation coefficient between X_i and X_j by the Bravais-Pearson formula.

By having a sample size n :

$$\rho = \frac{\text{Cov}(X_i, X_j)}{\sqrt{\text{Var}(X_i)\text{Var}(X_j)}}$$

$(x_i^1; x_j^1); (x_i^2; x_j^2); (x_i^n; x_j^n)$ from a joint distribution, the amount:

$$r = \frac{\sum_k (X_i^k - \bar{X}_i)(X_j^k - \bar{X}_j)}{\sqrt{\sum_k (X_i^k - \bar{X}_i)^2} \sqrt{\sum_k (X_j^k - \bar{X}_j)^2}}$$

Pearson's correlation coefficient is the covariance of the two variables divided by the product of their standard deviations. Therefore and taking into consideration the parameters related to the learners behavior inside the system, the covariance is as follows:

$$\begin{aligned} \text{Cov}_{x,y} = & ((\text{NBREXR}_x - m_x)(\text{NBREXR}_y - m_y) \\ & + (\text{NBREXM}_x - m_x)(\text{NBREXM}_y - m_y) \\ & + (\text{NBRAST}_x - m_x)(\text{NBRAST}_y - m_y) \\ & + (\text{ORD}_x - m_x)(\text{ORD}_y - m_y) + (\text{TMPTH}_x \\ & - m_x)(\text{TMPTH}_y - m_y) + (\text{FC}_x - m_x)(\text{FC}_y \\ & - m_y) + (\text{TCE}_x - m_x)(\text{TCE}_y - m_y) + (\text{DP}_x \\ & - m_x)(\text{DP}_y - m_y))/8 \end{aligned}$$

Where the standard deviation of each learner is:

$$\begin{aligned} S_x = & ((\text{NBREXR}_x - m_x)^2 + (\text{NBREXM}_x - m_x)^2 \\ & + (\text{NBRAST}_x - m_x)^2 + (\text{ORD}_x - m_x)^2 \\ & + (\text{TMPTH}_x - m_x)^2 + (\text{FC}_x - m_x)^2 + (\text{TCE}_x - m_x)^2 \\ & + (\text{DP}_x - m_x)^2)/8 \quad S_y = ((\text{NBREXR}_y - m_y)^2 \\ & + (\text{NBREXM}_y - m_y)^2 + (\text{NBRAST}_y - m_y)^2 \\ & + (\text{ORD}_y - m_y)^2 + (\text{TMPTH}_y - m_y)^2 + (\text{FC}_y - m_y)^2 \\ & + (\text{TCE}_y - m_y)^2 + (\text{DP}_y - m_y)^2)/8 \end{aligned}$$

If ($\rho \geq 0,5$)

1. Recommend the versions of the learning objects of this specific learner (with whom the similarity is optimum)

If ($\rho \leq 0,5$)

1. Search the similarity with all the Learners.
2. Update the profile by editing the learning style of the struggling learner according to the similarity result.

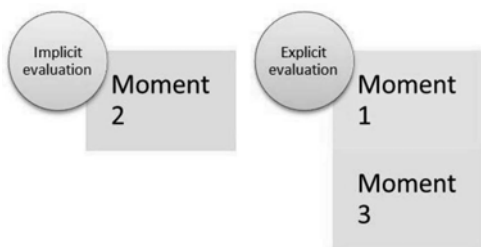


Figure 13. The correlation between the moments and types of evaluation.

4.3 *Moment 3*

The Third evaluation moment is a summative evaluation which aims to identify learners acquisitions and certify their skills; it represents a barrier to the passage to the next learning objects. Obviously, the moments 1 and 3 represent an explicit evaluation moment, where the moment 2 evaluation stands for an implicit evaluation moment. This situation is explained in the figure below.

5 CONCLUSION

Through this paper, we put the focus on the different aspects of the evaluation process, as it remains the only way to validate a learning process, then we high-lighted the crucial moments where it should happen in an adaptive learning system which takes into consideration the correction of the non-leading learning paths. Finally we intend to test the efficiency of the evaluation right on the moments revealed in this paper, using the ALS_CORR[LP] and a C programming language course as a starting point, the results of this experiment will be discussed in future works.

REFERENCES

- Bontchev, B. & D. Vassileva (2012). Courseware adaptation to learning styles and knowledge level. Edited by Anderson Silva, Elvis Pontes, I.
- Bousbia, N., I. Rebaï, J.-M. Labat, & A. Balla (2010). Analysing the relationship between learning styles and navigation behaviour in web-based educational system. *Knowledge Management & E-Learning: An International Journal (KM & EL)* 2(4), 400–421.
- Carchiolo, V., A. Longheu, M. Malgeri, & G. Mangioni (2007). An architecture to support adaptive e-learning. *International Journal of Computer Science and Network Security* 7(1), 166–178.
- Felder, R. M. & L. K. Silverman (1988). Learning and teaching styles in engineering education. *Engineering education* 78(7), 674–681.
- Felder, R. M. & J. Spurlin (2005). Applications, reliability and validity of the index of learning styles. *International journal of engineering education* 21(1), 103–112.
- Graf, S. & K. Kinshuk (2007). Providing adaptive courses in learning management systems with respect to learning styles. In *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, Volume 2007, pp. 2576–2583.
- Graf, S., S. R. Viola, T. Leo, & Kinshuk (2007). In-depth analysis of the felder-silverman learning style dimensions. *Journal of Research on Technology in Education* 40(1), 79–93.
- Guerrero-Roldán, A.-E. & J. M. Alfonso (2007). Adaptive learning paths for improving lifelong learning experiences.
- Jean-Daubias, S. Thi-thu-hong phan. Different levels of modeling for learner profiles.
- Nunan, D. & C. Lamb (1996). *The self-directed teacher: Managing the learning process*. Cambridge University Press.
- Papanikolaou, K. A., M. Grigoriadou, H. Kornilakis, & G. D. Magoulas (2003). Personalizing the interaction in a webbased educational hypermedia system: the case of inspire. *User modeling and user-adapted interaction* 13(3), 213–267.
- Seghroucheni, Y. Z., M. A. Achhab, & B. E. E. Mohajir (2015). Implementation of an adaptive learning system that include correction of learning path based on the differentiated pedagogy and the bayesian network. *iJES* 3(2), 27–31.
- Seghroucheni, Y. Z., B. E. E. Mohajir, et al. (2014). Exploitation of the recommendation systems in the calculation of the learning path. In *2014 5th International Conference on Information and Communication Systems (ICICS)*.
- Service Oriented Approaches and Lifelong Competence Development Infrastructures, 137.
- Tzouveli, P., P. Mylonas, & S. Kollias (2008). An intelligent elearning system based on learner profiling and learning resources adaptation. *Computers & Education* 51(1), 224–238.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Towards a blended learning using mobile devices, podcasts and QR codes in Algeria

Soulimane Ghizlene, Kouninef Belkacem & Djelti Mohamed

*Institut National des Télécommunications et des Technologies de l'Information et de la Communication,
University of Oran 1, Ahmed Benbella, Oran, M'Naouer, Algérie*

ABSTRACT: Higher education in Algeria has witnessed significant reforms in its educational system with a growing number of students from year to year due to its young population and a dynamic transition in the integration of Information and Communication Technologies (ICTs).

Algeria is gradually advancing in terms of telecoms and Internet. The fixed network is difficult to access. With mobile operators, Algeria is distinguished through the use of Mobile. In Algeria the mobile penetration rate stands at over 111% and 21% with 3G.

Internet access remains inaccessible to our students. But almost all students have access to mobile technology.

Indeed, learners now want to learn “on the move”. They are nomadic learners who learn in faculty, restaurant, library, before sleeping, around a coffee. But they also learn in communities (social networks) i.e they exchange them with unprecedented ease of information solutions to the problems and mutually explain what the professor said.

With this Internet generation, we have to change our methods of teaching and learning to think fast and efficiently, with a minimum of organizational, logistical and above all loss of time.

Mobile technology is increasingly being used to support blended learning. The satisfactory results of our previous research show that the use of mobile technology could enhance accessibility and communication in a blended learning course.

We discuss the emerging trends that allow more involved learners, such as e-learning, m-learning, blended learning, the use of podcasts and using QR codes in Learning Management System (LMS) at the National Institute of Telecommunications and ICT (INTTIC). QR Code is still relatively new and still in its infancy in education. The use of QR codes in INTTIC-LMS can be placed in the context of mobile learning.

Keywords: blended learning, mobile learning, Learning Management System, podcasting, QR code, Moodle platform

1 INTRODUCTION

M-learning is a term used to describe any manner of delivering courses or consultations of events spread through e-learning using mobile devices such as Pocket PCs, mobilephones or PDAs.

Mobile learning is a learning combining e-learning and mobile computers, and it allows the learners understand and enjoy education training at any place and any time through mobile facilities. One reason that m-learning has become so popular recently is the rise of ‘mobile Internet’. It is now common for many people to access YouTube, videos, their email and surf the web on their smartphones. All the interest of mobile media is the rapid diffusion of short information concerning training management: Advertisements or announcements on mobilephones (e.g. change of schedule, classrooms, meetings, news on forums, etc...)

Using available tools “mobiles” for education may seem uninteresting at first glance. All course contents cannot therefore be displayed on such tools. So, we can imagine contents of a “light type”. For example, “News”: results to show for an examination or other events, “notes” or the opposite, sending SMS and/or email from the platform to a PDA or a PC, etc.

Since most of our students have access to mobile technology, three in five were smartphones, this number that is expected to climb to over 65% by the end of 2015. Using this technology would encourage students to use their phones to send questions to their teachers, see the platform, ads, and grades, listen a podcast and snip the Quick Response (QR) codes. We propose some aspects of educational innovations. It is integrating rapidly evolving modes of learning impacted by the technological revolution. We carried out the implementation of

vCard, Quizzes and QR voice as a new tool in the Moodle platform. The QR code contains the URL of the page on one particular Moodle course and quizzes are added to the bottom of Moodle. Students' satisfaction had been acknowledged as an important factor in order to estimate the effectiveness of a blended learning course.

2 INTTIC EXPERIENCE

In Algeria INTTIC created since 1971 under the tutelage of both MPTIC (Ministry of Post Telecommunication and Information Technology Communication) and MESRS (Ministry of Higher Education and Scientific Research).

INTTIC is the only institution specialized in training in telecommunication and ICT in Algeria.

Furthermore and since its creation the INTTIC has trained 100 Magister (post-graduation), 2250 engineers, 4000 senior technicians, 13997 trainees in continuing education and 112 Certified Cisco Network Associate (CCNA 1, 2, 3, 4).

The INTTIC acquired its experience in e-learning when collaborating in the development of the INTTIC LMS e-learning platform with the University of Nantes (France), implementing and managing this platform since 2006.

Since January 2010, our e-learning research team took an initiative to use Moodle, for its variety of tools and good accessibility (Rice W. and S.S. Nash 2010). The platform is shown in Figure 1.

According to the new objectives of our team, we want to create an e-learning system that responds to three major axes of interest: (Rice W. and S.S. Nash 2010).

a. *Blended learning*: This learning solution is claimed to be “the most prominent instructional delivery solution”. In our case, it is mainly directed to students who attend courses at the

institute and need additional materials and skills. The combination of face-to-face learning with typically web-based educational technologies can enhance the quality of teaching and learning with lower price and less human resources. The proposed solution uses both asynchronous media like email, forums, weblogs or wiki in conjunction and synchronous media like text, chat or video conferencing.

b. *Distance learning*: The audience targeted by this type of education is employees from companies and institutions who need certification in certain courses (IT technologies, computing, networking...). These trainees are not (most of time) present on site, so the e-learning center must provide access to learning when the source of information and the learners are separated by time or distance, or both.

c. *Continuing education*: It is post-compulsory education (in addition to that received at secondary school), that is distinct from the education offered in tertiary education. This type of training is dedicated to professional companies to improve capacity building of their personnel in related technological fields.

Let us take a closer look now at Blended and Mobile learning:

2.1 Blended learning

Blended learning has been defined in complex ways but generally assumes a combination of real time and online interaction, often through the medium of integrated learning management systems. The concept of blended learning is defined in a variety of ways with different dimensions of the blend identified by (Singh H., Reed C. 205), among others. (Stacey E. 2015) recognized that a blended learning program may combine one or more than six dimensions: offline and online learning; self paced, live and collaborative learning; structured and unstructured learning; custom content with off-the-shelf content; work and learning; and ingredients of the blend: Synchronous physical formats, synchronous online formats, and self paced, asynchronous formats.

It is a fact that the success of distance education depends largely on student support services provided to its learners who encounter a feeling of isolation, lack of peer-peer interaction, lack of proper intimation from study center, lack of proper academic support and hurdle of distance from the study center to list a few (Lalita S. Kamar and al. 2011).

The platform could be used at home, but also during the course. Our approach aims to redefine the role of the teacher while reaffirming its central



Figure 1. Module created in INTTIC-Moodle.

place in the device as a designer and students tutor.

The platform provides online resources (texts, sounds, videos, Power Point presentations etc.) and activities (deposit files, chat, forums, interactive quizzes, glossaries and surveys).

Previous experiments using blended learning show satisfactory results in improving students' competences (Alsaffar A.A. and E. Namkeh 2011).

2.2 M-learning approach

2.2.1 Applications

The use of the current "mobile" tools in the teaching profession can appear without interest. The connection speed on mobile devices can be slow. Using a small hand-held device is not really ideal for accessing, say, distance course materials on Moodle. One obviously thinks to read courses on as a small tool as phones or PDA and one includes/understands the difficulty quickly. All course contents could not be posted on this kind of tools. One can thus imagine light contents of a "summarized" type, "the points to be seen for an examination", etc.

Although these tools are becoming largely used, our problem remains actuality, the screens size remains too small for the reading. Even though the battery autonomy is in constant improvement, their use remains weak compared to the awaited uses for the e-learning. The user interfaces are not convivial on the majority of mobile phones and finally the diversity of the mobile apparatuses and the tendencies fast change insist to think of producing "mobiles LMS" able to adapt to a whole mobile line of goods rather broad.

The following applications seem to be of interest to the students:

Current events consultation in INTTIC-Moodle: This application makes it possible to students to quickly consult short information concerning current events management on the mobile phones.

- *Grades results consultation:* This application concerning grades management makes it possible for students to consult their grades starting from their mobile phones. There are two consultation possibilities: by module and/or by education level.
- *SMS/Email connection:* Means sending from LMS towards the mobile. The connection costs still remain high. However the work advantages on the mobile are numerous and with media use, the training concept at anytime and anywhere will be real. It will be based on the new society practices which like to make profitable empty times, which appreciate to consult by short moments to avoid displacement. Since the mobile use, it is possible to get informed about the current events (Vanaja M. 2010).



Figure 2. Current events consultation.



Figure 3. Grades results consultation.

Internet is not accessible for all students, but almost students have access to mobile technology. Short Message Service (SMS) means exactly what its words suggest; it is a message sent to students. The use of SMS is becoming widely implemented in education. Wireless is a particularly attractive option for blended learning.

2.2.2 Podcasting process

Using this process, the teacher will be able to prepare fast videos in order to facilitate the learning process using podcasting as a means of sound files diffusion (Motiwalla L.F. 2007).

Listening to numerical audio contents will not easily replace the reading, listening on line presentations or any other means of receiving information has to learn [4].

Nevertheless, here are some points which show podcasting contribution in teaching (Nataatmadja L. and L.E. Dyson 2008).

- Some learners prefer to obtain oral information: the podcasting would then make it possible to provide learning supports from courses (audio) which are adapted to them.
- To provide the students with additional course notes at lectures time, student who are not able to follow the teacher presentation, take a notes and related explanation, podcasting could be used as an oral additional tool for students.



Figure 4. SMS/Email.



Figure 5. Podcasting tool.

- Podcasting could be a good help and assistance to those who have difficulties understanding face to face lectures. The use of podcasting can then be an excellent means for course revision and help the teachers improve their presentation manner has the oral examination course.

3 INTRODUCTION OF QR CODE IN MOODLE

A QR Code is a Matrix code (or 2D bar code) developed by a Japanese company DENSO WAVE in 1994 (Denso 2013). QR stands for 'Quick Response' as a barcode readable by a mobile device with a camera. Information such as multilingual text, a linked URL, SMS message, or other data can be easily encoded using a QR code generator. QR codes are capable of encoding any type of data (7089 numeric characters), (4296 alphanumeric characters), (2953 binary bytes) and (1817 Kanji characters).

A QR code has the capacity to store its information horizontally and vertically indeed the reading will be done on two axes. A QR code can be represented by a matrix (x, y) Figure 6.

Technology progress in the field of learning has evolved especially with the use of m-learning and now with the emergence of QR codes. QR codes are still relatively new and not widely known in education (Dragulescu B. 2012). Many scholars see the potential for using QR codes to direct students to RSS feeds, lecture podcasts and other just-in-time



Figure 6. Example of QR code.

resources. We carried out some examples of applying QR codes in INTTIC-LMS.

3.1 Using vCard and QR code in Moodle

As presented in the introduction, in Moodle the process of saving the teacher's personal data into an electronic agenda is rather difficult. First the user has to click on the teacher's name, and then he is redirected to a page with the required information, which he has to manually insert into the application of choice.

The QR code can be read by a mobile phone with a photo camera and a QR reader installed. The result is a business card of the tutor, containing the information retrieved from the Moodle database, information which can be easily saved in the telephone's contacts list.

We had two possibilities to accomplish the task mentioned above: The vCard generated by our application could be encoded directly into the QR image, or we could encode a link to the vCard (Figure 7).

3.2 Generating the QR codes

Generating QR codes has been made simple because there are many free QR code generators on the Internet. QR codes are added to the bottom of Moodle. The QR code contains the URL of the page (or text) of the Quiz (Figure 8).

A student snipped the QR Code on the Quiz is shown in Figure 9.

Each of the QR codes is a text file. There is no need for the mobile devices to connect to the Internet to decode them. The correct answers for this quiz are shown in Figure 10.

In addition to the podcasts we can generate a code with a message in almost all languages. QR voice is a code with a text to speech application hidden in it (Figure 11). The QR voice turns the



Figure 7. Teacher's personal data with QR code.



Figure 10. Correction of quiz.

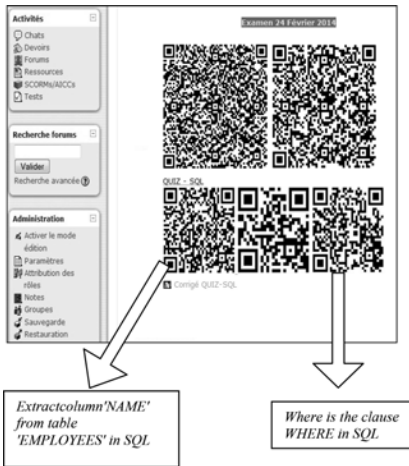


Figure 8. Example of quiz.



Figure 11. Example of QR voice.



Figure 9. A student scanned the QR code on the quiz.

text in an audio. Once scanned it will play the message with a synthesized voice using the translation function Google with Voice integration of a robot that lets you listen to the translated text (Xiaoji Jiang et al. 2013).

The icon representing an earphone allows you to hear the message.

4 RESULTS AND DISCUSSIONS

The structure and the activities composing the system are the result of two years of experimentation during which many improvements were produced.

By looking at the two experiments, we propose the use of two courses in the Moodle platform.

4.1 Mobile process

The self-reported number of times that respondents indicated about sending texts is contrasted with the actual recorded number of text messages. The rate of using SMS and the number of students over a period of time is shown in Table 1.

4.2 Podcasting process

In Table 2, we range a number of different m-learning activities in three months in the second year of use with podcasting approach. From this work for considering m-learning, we can see that podcasting alone will not improve education since it largely perpetuates the traditional teaching.

This study also involves the use of QR codes and mobile devices which may help advance the m-learning at INTTIC (Kouninef et al. 2012). Another advantage of using personal mobile devices is that the institution need not devote time and resources to procure devices when most learners already own devices that support m-learning. As has been said earlier, although ownership amongst the learners who participated in this study is average, national and global trends in the telecommunication industry indicate that this is likely to grow within the next few years. The use of QR codes eliminates the irritating issue of keyboard input.

There is, however, a small percentage of learners who were dissatisfied, and the reasons are mainly due to the fact that in the first place, they do not own a smartphone or are reluctant to pay for the Internet charges. They suggested that Algeria Telecom bears some of these costs if the INTTIC is to encourage a more pervasive use of such a system. Additionally, a few complained of a lack of

Internet access, making it difficult for them to make full use of the QR codes.

It is also gratifying to note that learners to continue using QR codes was above average, with satisfaction being the most significant factor that impact the learners to continue using QR codes. The use of QR codes in LMS is expected to compel learners to fully utilize these readily available resources.

Over the 60% of the students are satisfied with blended courses.

The results support the conclusion that students' satisfaction was higher than average.

The survey revealed that 61% of the 100 learners who participated in the survey owned a smartphone. These figures show that accessing learning materials via smartphones or tablet computers is currently universally applicable to all learners in the INTTIC. While the survey found that 58% of learners have scanned QR code to access the questionnaire other learners have used the URL (Figure 12).

It was found that the learners were moderately efficient in using computer technology and that satisfaction has significant impact on learners to continue using QR codes. As the study is based on a small sample of 100 students, the results obtained should be interpreted with caution. Nevertheless, these results provide some aspects that influence learners to use QR codes.

The learners were asked if they were satisfied with the use of QR codes for learning. The majorities were satisfied; and explained that the system was easy to use; learning with QR codes was fun; and the system also renders greater mobility in learning.

There are, however, some learners who had some reservations about this approach. In the first place, these learners do not own smartphones or are reluctant to pay for the required Internet charges. They suggested that Algeria Telecom subsidizes

Table 1. Mobile phone use.

Duration	Rate of sending SMS	Number of students
One year	51%	105
2 years	80%	183

Table 2. Podcasting use.

Podcasting	Using time (hours)
Anywhere anytime access	120
Interactive classroom	35
Mobile phone communication	88
Multimedia data capture	24

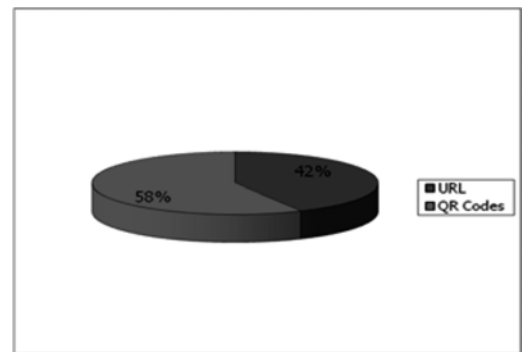


Figure 12. Type of access to the survey.

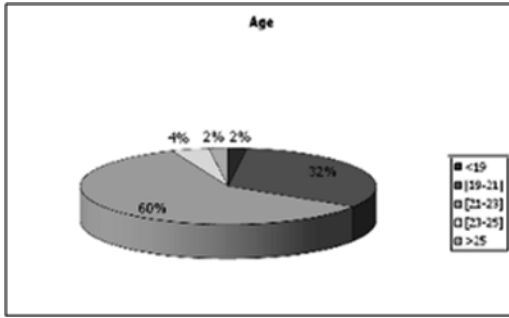


Figure 13. Student participate to the study.

Internet charges if the INTTIC is to use QR codes. A few also complained that they were not able to access the Internet and hence, could not access the embedded information in the given QR codes.

Some also asserted that the smartphone might not be suitable as a device for learning as its screen is rather small and its battery life is limited. Learners also suggested that INTTIC provides better WiFi for easy access to QR codes.

The effectiveness of blended learning instruction was investigated by looking at the students' satisfaction. A sample of one hundred ($n = 100$) undergraduate students between the ages from 19–25 years old participated in the study (see Figure 10). Forty-two percent (42%) of the participants were female and fifty-eight (58%) were male. For the data collection at the end of this investigation, students completed a questionnaire including the students' demographic/personal data and evaluated their students' satisfaction about the use of blended learning courses.

With the development of new technology, we have added the ability to capture student feedback with QR codes. A QR code can be added at the end of courses in the LMS. By scanning the code with their smart phones, students can access the survey on student satisfaction. This survey is programmed on our platform, to give students an easier way and more convenient to express opinions and to offer an innovative way to access student feedback.

5 CONCLUSION

In a context marked by communication and information technologies development used in education and especially applications development uses mobile technologies. We are witnessing the emergence of the e-learning in parallel to m-learning where the coexistence of these two environments makes it possible to develop learning independ-

ently from the time and place restrictions. Mobility becomes a data processing key factor (Mamlook R. et al. 2011). The mobile phones become genuine small computers and offer still under-exploited capacities. In this paper, we tried to take account of the specificity of the m-learning so that the students of owner institute can profit from this training in technology and to place at their disposal a mobile platform Learning Management System. Through this system, the teacher can prepare audio and video sequences for his/her students downloadable from their iPods. This is especially useful for foreign students who suffer from language difficulties.

We provided also a view of using mobile technology and QR codes at INTTIC. We generated vCards for fast and easy access of the contact information of teacher. We used QR codes for mobiles, so that vCards can be imported into smartphones.

We are witnessing the emergence of the e-learning in parallel to m-learning where the coexistence of these two environments makes it possible to develop learning independently from the time and place restrictions. Mobile and online technology has enormous potential to transform learning [7]. We tried to take account of the specificity of the m-learning so that the students of owner institute can profit from this training technology and to place at their disposal a mobile platform Learning Management System.

In general, we believe that QR codes have great potential in Higher Education. Some possibilities are shown in this paper. Mobile devices using QR codes at INTTIC are playing an increasingly important role in this ever changing learning environment. Through this system, the teacher can prepare QR voice for his/her students downloadable from their smartphones. This is especially useful for foreign students who suffer from language difficulties. It is gratifying to know that learners appear to be positive on a platform that uses QR codes, smartphones and tablet computers. In future work, we want to exploit all available tools and features in Moodle and further integrate other types of QR codes.

Finally, it may be a good idea for INTTIC to partner with telecommunications operators (Algeria Telecom, Mobilis, Ooredoo, Djezzy etc.) in subsidizing the costs of purchasing mobile devices and Internet charges.

In future work, we want to exploit all available tools and features in Moodle and to profit from its technical advantages.

ACKNOWLEDGEMENTS

Authors are grateful to respondents of this study.

REFERENCES

- Alsaffar A.A. and E. Namheh, "Secure Migration Service for Mobile IPTV Using DCAS", *Information Technology Journal*, Vol. 10, N°11, pp. 2044–2051, 2011.
- Denso "Standardization retrieved Juin 2013 from <http://www.denso-wave.com/qrcode/aboutqr-e.html>.
- Dragulescu B., Ermalai I., Bucos M., Vasiu R., "Metadata Methods for Improving Usability in Moodle" *International Journal of Web Engineering* 2012, 1(1): 6–10.
- Katz J.E. and M.A. Aakhus, *Perpetual contact: Mobile communication, private talk, public performance*. Cambridge UnivPr, 2002.
- Kouninef B., Tlemsani R., Rerbal S.M., Lotfi A. "Developing a Mobile Learning Approach in Platform LMSINTTIC", *Information Technology Journal* Volume 11, Number 8, 1131–1137, 2012.
- Lalita S. Kamar and al. "Mobile device Intervention for student support services in distance education context-FRAME model perspectives" *Eurodl* 2011.
- Mamlook R., A. Aljumah and N.K. Farooqui, "Knowledge on the Move", *Journal of Applied Sciences*, Vol. 11, N°16, pp. 3062–3069, 2011.
- Motiwalla L.F., "Mobile learning: A framework and evaluation," *Computers & Education*, Vol. 49, N°3, pp. 581–596, 2007.
- Nataatmadja L and L. E. Dyson, "The role of podcasts in students' learning," *iJIM*, Vol. 2, N°3, pp. 17–21, 2008.
- Rice W. and S.S. Nash, "Moodle 1.9 teaching techniques," 2010.
- Singh H., Reed C., "Achieving Success with blended Learning—Leerbeleving, Centra Software, <http://www.leerbeleving.nl/wbt2014/blend-ce.pdf>. [Viewed October, 22, 2015].
- Stacey E. "Effective Blended learning Practices-<https://books.google.fr/books?isbn=1605662976>, [Viewed October, 01, 2015].
- Vanaja M., "SMS Advertisement: Competitive to Gulf Market?", *Asian Journal of Marketing*, Vol. 4, N°3, pp. 131–143, 2010.
- Xiaoyi Jiang, Matthew Ma, Chang Chen, Tsung-Yu Liu, Tan-Hsu Tan, Yu-Ling Chu, "QR Code and Augmented Reality-Supported Mobile English Learning System", in *Mobile Multimedia Processing*, vol. 5960: Springer Berlin / Heidelberg, pp. 37–52.

Evolutionary algorithm to solver impairment aware wavelength assignment problem

A.M.L. Miranda & C.A.J. Rocha

Coordinating of Informatic, Federal Institute of Pará, Pará, Brazil

J.C.W.A. Costa & C.W.A. Costa

Institute of Technology, Federal University of Pará, Pará, Brazil

ABSTRACT: In this paper, we propose a hybrid methodology based on Graph-Coloring and Genetic Algorithm to solve the Impairment Aware Wavelength Assignment problem. Our proposal was developed for a static scenario, and shows a significant reduction in blocking probability.

1 INTRODUCTION

Optical networks based on Wavelength Division Multiplexed (WDM) are widely deployed in backbone networks due to their large bandwidth available over long distances. However, finding an optical solution for Routing and Wavelength Assignment (RWA) algorithm in the design and operation of the networks remains an open issue. A special class of RWA algorithm is the one considering accumulated Physical Layer Impairments (PLIs), and is known as Impairment Aware Routing and Assignment Wavelength (IA-RWA) algorithm, Jirattigalachote (2012).

These PLIs are classified in two categories: linear and nonlinear impairments. Accumulation of linear impairments is directly proportional to the length of the link, Monoyios (2009), hence, choosing the shortest route can reduce them. Conversely, the most important nonlinear impairments are: Self-Phase Modulation (SPM), Cross-Phase Modulation (XPM) and Four-Wave Mixing (FWM). However, Ten et al. Ten (1999) compared FWM and XPM penalties for a 40 x 10 Gb/s system with channel spacing of 100 GHz where they found that the degradation of the optical signal affected by XPM is several times greater than that of FWM. In the scenario presented in our paper, the dominant nonlinear impact is the XPM effect.

Our proposal is to be used in the design phase of the network (static scenario), as long as the physical topology and traffic matrix are known a priori. For the routing sub-problem, we used the Dijkstra algorithm to find the shortest-path. We chose this algorithm because a shorter route possibly reduces some linear impairments (e.g., ASE noise and chromatic dispersion). With respect to Wavelength Assignment (WA) sub problem, we created a meth-

odology based on the use of metaheuristic. Our work minimizes the total number of wavelengths required for the network and determines the wavelength activation order, reducing the XPM effect.

2 CROSS-PHASE MODULATION MODELING

Cross-phase modulation is a nonlinear phenomenon that occurs when two or more optical channels having different wavelengths propagate simultaneously inside an optical fiber. The phase of each channel is modulated by the Intensity Modulation (IM) of the other channels.

The analytic model developed by Cartaxo et al. Luis, R. S. (1999) analyzes the XPM effect in WDM systems based on intensity modulation and phase modulation. They are at the output of the transmission system of the analyzed channel, i.e., probe channel and are caused by one or more interfering channels pump channels.

In order to estimate the quality of the transmitted signal, the normalized variance of XPM-induced IM is employed. Normalized variance of XPM effect can be written as the integral of the Power Spectral Density (PSD) of the probe channel IM, normalized by the square of the power at the symbol 1, as follows, Luis, R. S. (1999):

$$\sigma_n^2 = \frac{1}{P_i} \sum_{j=1}^M \int_{-\infty}^{+\infty} S_{p,j}(f) \cdot |H_{XPM,p,j}(f)|^2 \cdot |H_i(f)|^2 \cdot df \quad (1)$$

Where \bar{P}_i is the average optical power of the probe channel, M is the number of interfering channels,

$S_{p_j}(f)$ is the power spectral density of the j 'th pump channel IM at the fiber input, H_{XPM,p_j} and $H_r(f)$ are the transfer functions of the equivalent linear model of the XPM-induced IM associated with the j 'th pump channel and of the electrical receiving filter, respectively. The expressions for each of the described elements of Eq. (1), implemented in our paper, can be found in Cartaxo et al. Luis, R.S. (1999). The analytical model for characterizing XPM in multi-span optical systems used in this article is similar to that used in Cartaxo et al. Luis, R.S. (1999).

3 PROPOSED METHODOLOGY

Fig. 1 shows the hybrid methodology proposed which is composed by a graph coloring algorithm, followed by a Genetic Algorithm (GA). The graph coloring algorithm is initiated immediately once the routes, that will meet the traffic matrix, are defined. The goal is to assist the wavelength assignment process by minimizing the number of necessary wavelengths, since the reduction in the number of wavelengths is an effective way to minimize the effect of XPM. With this number, the active wavelengths on the wavelengths grid must be spread, fact that minimizes the influence among channels.

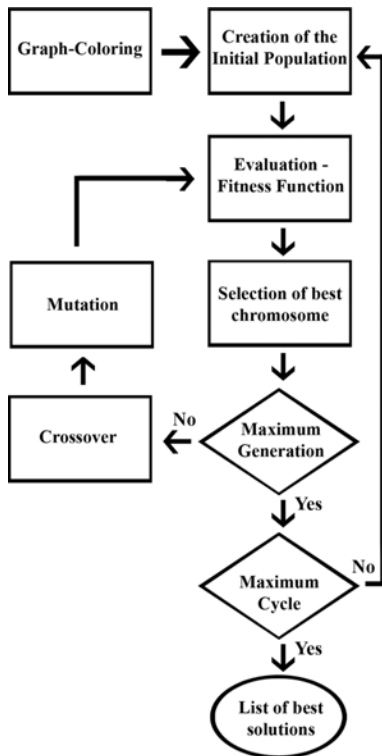


Figure 1. Flowchart of proposed methodology.

To find optimal or near-optimal solutions to spread the wavelengths on the wavelengths grid, we use a GA. Our proposal performs an optimization iteratively evaluating the various candidate solutions in large-scale search spaces, and the output is a list of the best solutions. GA starts creating an initial population, where each chromosome is randomly generated and represents a codified version of the order of activation of wavelength in the grid wavelengths, and then the following rules apply, at each generation:

- Fitness function—composed by a formulation that calculates the normalized variance of XPM effect, expressed in (1)
- Selection—selects the fittest chromosomes to produce better offspring
- Crossover and Mutation—genetic operators responsible for generating new chromosomes, allowing exploration of the search space of problem solutions.

When the generation number reaches the maximum value (20 generations), a new cycle starts with a new initial population and the process, of evaluation, selection, crossover and mutation, is repeated for 20 more generations. The number of cycles has been fixed to 30 to enable expanding the search space in different regions, thus increasing the chances of finding optimal solutions. At the end of each cycle, the best solution found is stored so that the final result ends up in a list with the best solution in each of the 30 cycles. To avoid premature convergence, we used the operations of crossover and mutation rate of 0.8 and 0.01, respectively.

4 SCENARIOS DESCRIPTION AND RESULTS

Our scenario considered each optical link of 40 wavelengths with data rate of 10 Gbps per channel. The simulations were performed in an optical network very close to the actual situation in terms of distance between nodes, which provides more realistic results. Fig. 2 shows the network topology, which has 16 nodes, 22 optical links (all links above 100 km were randomly divided in spans with size ranging from 50 to 100 km), and 302 passage nodes (nodes allocated in optical links, which may be used to add amplifiers and dispersion compensators, for creation of span).

The configuration model of the fiber link assumed in each span, composed by a sequence of alternating Single-Mode Fiber (SMF), Dispersion Compensation Fiber (DCF) and amplifier. Our tests were performed with the number of connection requests ranging from 10 to 40, in two different scenarios, S1 and S2. S1 scenario considered blocked connections

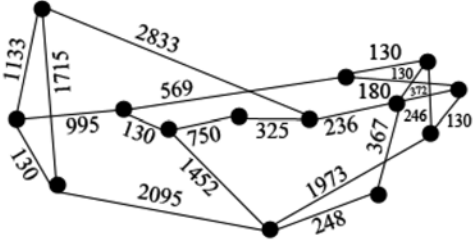


Figure 2. NSFNET network topology (distance in km).

Table 1. Default simulation parameters.

Parameter	Values
WDM grid spacing	100 GHz
Lowest wavelengths of the grid	1529.55 nm
Data rate per channel	10 Gbps
Amplifier noise figure	5.5 dB
SMF loss coefficient	0.22 dB/km
SMF dispersion coefficient for 1550.12 nm	17 ps/km.nm
Dispersion slope of the SMF	0.08 ps/km.nm ²
DCF loss coefficient	0.5 dB/km
DCF Dispersion coefficient	-100 ps/km.nm
Dispersion slope of the DCF	-0.3 ps/km.nm ²
Fiber nonlinear coefficient	1.37 (W.km) ⁻¹

only by XPM effect, ignoring blocked lightpaths by residual dispersion, while S2 scenario considered blocked connections by both XPM and residual dispersion effects. In the latter scenario we considered blocked all lightpaths with residual dispersion values from 1175 ps/nm based on ITU, Geneva (2009). In both the scenarios we used the parameters listed in Table 1. The optical channels in a determined span have equal power values, but these values can be adjusted from -2 to 8 dBm depending on span length, in order to guarantee a convenient power budget. Figures 3 and 4 compare the blocking probability in our proposal with the First-Fit algorithm as a function of the number of connection requests for two scenarios. As can be seen, the GA shows an improvement of rate of blocking probability. This is expected since the GA tends to spread the wavelengths on the wavelengths grid, fact that minimizes the influence among channels. Therefore, our proposal is efficient when the number of demands below 40, and we considered only the XPM effect. When we also consider the residual dispersion our proposal presents good results for maximum of 25 demands.

One can also notice that only by using preprocessing Graph-Coloring algorithm, it is possible to observe a significant improvement over traditional First-Fit algorithm, due to reduction the number

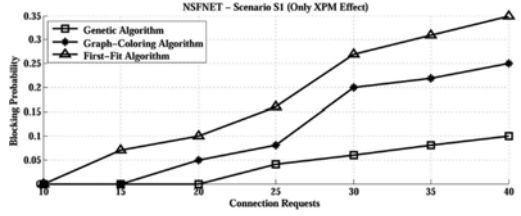


Figure 3. Blocking probability in S1 scenario.

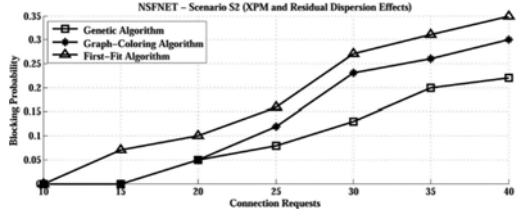


Figure 4. Blocking probability in S2 scenario.

of wavelengths (channels) needed to meet traffic matrix. Analyzing Fig. 4, one can see that in the case of the First-Fit algorithm the blocking probability reaches 7% for a demand of 15, which is attended by 4 wavelengths. While GA meets almost double (25 demands) with nearly the same blocking probability, but with only 3 wavelengths. Thus, one can notice the reduction not only in the cost with equipment but also with the network energy consumption.

The execution time grew exponentially according to the number of demands required. In order to perform 10 demands the optimization process took around 22 minutes, around 1 hour for 25 demands, and around 3 hours for 40 demands.

5 CONCLUSIONS

The results for GA showed the creation of multiple lists containing the best solutions in terms of blocking probability. Each list created in offline phase was generated by a different demand matrix and could be used in the network operation phase according to the current network status. In all cases, it was found that the list generated by GA presented better results than the First-Fit and the Graph-Coloring algorithms. However in some situations the Graph-Coloring provides satisfactory results without the need to perform GA. This optimization process is flexible and in future it will be extended to incorporate other physical layer impairments, making the GA more efficient. This work demonstrated the applicability of computational intelligence and bio-inspired algorithms to solve different optimization problems involved in the design of networks.

ACKNOWLEDGEMENTS

We thank Federal Institute of Education, Science and Technology of Pará (IFPA).

REFERENCES

- Geneva (2009). Optical fibers, cables and systems. International Telecommunication Union (ITU).
- Jirattigalachote, A. (2012). Impairment-aware routing and waveband assignment for efficient optical transport networks. Optical Fiber Communication Conference and Exposition (OFC).
- Luis, R.S., C. (1999). Analytical characterization of spm impact on xpm-induced degradation in dispersion-compensated wdm systems. *Journal of Lightwave Technology* vol. 23, 1503–1513.
- Monoyios, D. (2009). On the use of multi-objective optimization algorithms for solving the impairment aware-rwa problem. *IEEE International Conference on Communications (ICC)*.
- Ten, S. (1999). Comparison of four-wave mixing and cross phase modulation penalties in dense wdm systems. *Optical Fiber Communication Conference and Exposition (OFC)* vol. 3, 43–45.

Using the 3M method for the optimization of the managerial act

Bogdan-Alexandru Furduescu

“Valahia” University of Targoviste, Targoviste, Romania

ABSTRACT: The conscious mind is that part of our mind in charge of logic and reasoning. Being dedicated to a certain task, the entire mental activity cannot offer the information that a conscious person can offer. Therefore, if the proper linguistic structures are applied, the mind cannot resist, even if a person might. In order to exert influence on the way a person thinks, the own semantic structures must be used, so that the communication barriers on the conscious level are broken. In order to achieve the proposed objectives, the *3M method* can be used within the organization, method which allows the use of three fundamental elements of NLP—the *Meta Model*, the *Metaphor* and the *Milton Model*—on optimum level, eliminating the elements generating resistance within the verbal and nonverbal communication process.

Keywords: meta model, metaphor, milton model, 3M method, Neuro-Linguistic Programming (NLP)

1 INTRODUCTION

NLP represents a way of opening new perspective which, through a wide range of practical methods and efficient techniques, instruments, procedures, models and theories of personal development with which cognitive and behavioral structures may be decoded, improving and developing any type of personality. “*The personal development theories are based on the postulate that people are animated by intrapsychic potentials which managers must locate and emphasize, offering concrete instruments directly applicable in managerial practice*” [15, p. 402]. From the theoretical and practical framework of three NLP fundamental elements – *Meta Model*, *Metaphor* and *Milton model* – the *3M method* is born, an unique concept with the purpose of not only improving the actions of these elements, but also a more efficient use of human resources.

The 3M method can be defined as a “*set of principles, rules and means of knowledge and transformation of reality which, in a purely physical sense, becomes a system of principles, rules and means of knowledge and transformation of reality*” [16, p. 137].

2 MATERIAL AND METHODS

The material published in this paper was elaborated after studying the theoretical aspects that form the basis of human resources management.

The methodology of elaborating this paper considers a complex analysis of the studies and theoretical aspects that form the basis of NLP. *The Meta Model*, the *Metaphor* and the *Milton model* may be considered methods because:

- *The meta model* uses certain instruments (called *operations*) in order to process the information collected within the organization in a certain way;
- *The metaphor* uses—mainly—the analogy, the symbolism in order to help find new solutions and to overcome certain communication barriers;
- *The Milton model* uses syntactic elements with the purpose to achieve the objective, but also pursues placing certain instructions (called *fixed orders*) in order to insert them and obtain maximum performances and efficiency.

3 RESULTS AND DISCUSSIONS

In the *Milton model*, the *3M method* aims to improve the technique generating a trance state so that the transmission of orders is facilitated, overcoming conscious barriers existing in the communication process, in the *meta model* being followed by the improvement of specific orders so that, using minimum resources, the two models give birth to an optimized flow of information. In the *metaphor*, the generated analogies are used in this method for various problems and optimum solutions for their solving.

4 META MODEL

This the first and most important model developed within NLP, being a set of linguistic models that reconstitutes the connection between deletion, deformation, generalization processes and the experience that gave birth to them. This *model* uses

language to highlight and overcome its limits and starts from analyzing the type of questions Frits S. Perls and Virginia Satir used in psychotherapeutic work, as well as some of the ideas emitted by Alfred H.S. Korzybski and contains two components [1, p. 189]:

- a theoretical vision on language;
- a series of questions meant to lead to the elimination of informational generalizations, distortions and shortcomings.

Richard W. Bandler and John T. Grinder combined them with the researches of Noam A. Chomsky about *transformational grammar* and they published the results in 1975 in the book “*The Structure of Magic. Volume I*”. The meta model represents the exposure of derivation of logic expressions and not expressions themselves. This model works on the conditions that words manage to create an *anchor* on the level of an individual through an experience or a sensorial representation. The *anchoring* is the process through which a stimulus or a representation (inner or outer) is associated with certain *external triggers* provoking a certain answer, which may be fast, or hidden, revisited. The NLP concept regarding *anchors* derives from the Pavlovian *stimulus* → *reaction* relation, which is a classic conditioning example.

Due to the fact that in Greek “*meta*” means “*with*”, “*after*” or “*near*”, the *meta model* reconnects the experience with the language, imposing itself as linguistic model about language, clarifying it through itself with the purpose of allowing a better understanding of the interlocutor and a better expression. We communicate with the help of words, resorting to deletion, deformation and generalization, principles we apply to the profound structure of our experience, in order to bring it to the stage of verbally expressed *surface structure* [14, p. 214]. Therefore, the meta model represents the linguistic differences through which one can identify language patterns that *distort*, *delete* or *generalize* significance in the communication process, allowing the partial *generalization* of information that the individual takes from the organization.

In order to fully understand the *meta model*, we must first analyze the way thoughts are transposed into words, the language not being able to keep pace – in any circumstances – with the speed of thought, with its variety and sensitivity, only being an approximation of it. The person has an idea, a full vision of what he/she wants to say, called *profound structure*. This structure is unconscious, the language being hidden on a very deep level of the human brain. In order to express himself/herself, the human being “*short-circuits*” the *profound structure*, what he/she pronounces being called *surface structure*. Both structures are necessary in

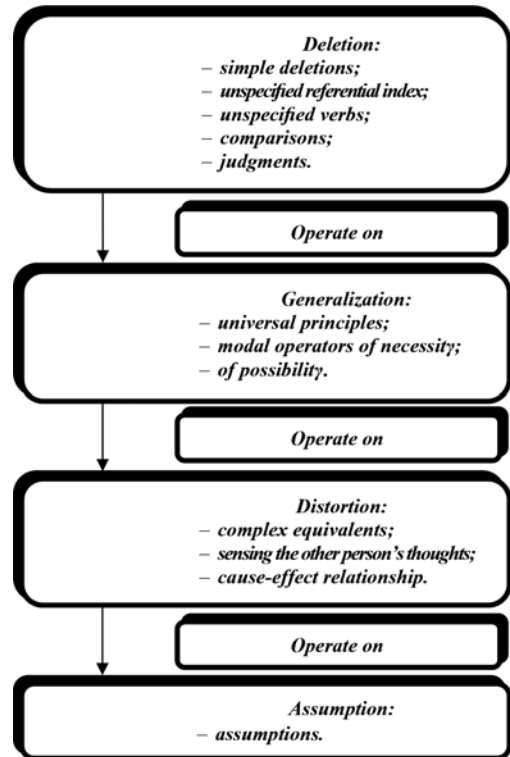


Figure 1. Operational levels of NLP meta model 9, p. 125.

different contexts. Questions regarding *meta model* produce a “*reverse engineering*” on language, in the meaning that it works with the *surface structure* for a better observation of the *profound* one behind it. In the transition from *profound structure* to *surface structure*, the subconscious activates the following operational categories:

- *Deletion* of a part of the information available in the *profound structure*, which limits thinking and acting. This category materializes in the following specific *patterns*: *simple deletions*, *unspecified referential index*, *unspecific verbs*, *comparisons* and *judgments*.
- *Generalization*, by assigning an universal value, based on a limited personal experience, all conditions and exceptions that would make a discussion boring or contradictory, which reduces the range of possibilities. In this category, the following specific *patterns* materialize: *universal principles*, *modal operators of necessity* and *possibility*.
- *Distortion*. a simplified vision of the information, which limits the interpretation options, modifies the meaning and leads to other useless problems. This category materializes in the

following specific *patterns*: *complex equivalents*, *sensing the other person's thoughts* and *the cause-effect relation*.

The three factors have a similar action: they take the *generalized* information or *distort* it in relation to a certain specific factor, a causal relationship, using a relationship between a linguistic structure and a certain environmental process, etc. This means that in the end the concerned person interpreted in his/her own way the information received from the environment [10, p. 46].

- *Assumption*. Ideas or assumptions considered as being *good* or *true* in order to make sense of the communication. In this category one single specific *pattern* materializes: *assumptions*.

It has the role of identifying missing information and replacing them so that, in the end, a certain result is produced. At the same time, *assumptions* have the role of determining, in the *cause-effect* relationship, the *complex equivalency* and *generalization*. The use of *assumption* can materialize in several directions, respectively: *the type of searched information*, *the minimum quantity of information for the respective individual*, *the feelings transmitted by the respective information*, *in relation to what objective is the concerned individual placed*.

Although the *Meta Model* uses *deletions* or *generalizations*, it has a fairly high degree of precision, as it uses the questions “*who?*”, “*what?*”, “*how?*” and “*in what way?*”; however, these are not used in case there is a lot of information, this leading in the end to the decrease of utility – overall – of NLP *meta model*. In order to achieve the *meta model*, it is necessary to take the following steps: *sorting different states* (achieved for *re-anchoring* different aspects of the information received from the organization), *accessing different states* (in order for certain *anchors* to be *revisited* and optimized) and *creating a meta part* (for the purpose of mutual conditioning between several pieces of information and assimilating an *anchor* new to them). Operations specific to *meta model* [10, p. 49–53]:

- *Deletion*—represents the elimination of that information deemed redundant or which cannot pass the *VAK sensory filter* (associated-dissociated meta method in each sensory representation system). Certain constructions such as “*I am happy!*”, “*I am uneasy!*”, “*I am confused!*” or “*I am scared!*” insure the recovery of certain previously *deleted* information.
- *Comparisons*—“*the best ...*” and “*the least good ...*” represent another form of deletion of information. “*The best ...*” or “*the worst ...*”, “*better than ...*” or “*worst than ...*”, “*in comparison to ...*” and such other comparisons are nothing but a form of *deletion*.

- *Unspecific indexed references* – are linguistic constructions such as “*which*”, “*what*”, “*these*”, “*people*” that have the feature of being able to *delete* information from sentences, but maintaining their sense.
- *Substantivizations*—insure the transformation of an action, usually a verb, into a noun, as static entity or object. *Substantivizations* refer to changing a process in the deep structure into a static event and they have the role of connecting with a certain part of the experience.
- *Unspecific verbs*—are, generally, sensory constructions that, at the moment of realizing a process, activate the senses, respectively sensory acuity. “*To believe*”, “*to know*”, “*to feel*”, “*to touch*”, generally the sensory verbs, are considered *unspecific verbs*. These *verbs* have the feature to make oneself change their mind in a certain process, but also to realize their senses.
- *Modal operators*—are of *possibilities* and of *necessities*. Aggressive persons tend to use *modal operations of possibilities*. In this category, a single specific *pattern* materializes: “*I can do everything!*”. *Modal operations of necessities* of a passive person show how he/she targets the fulfillment of own modalities, necessities, possibilities; we use the predicates “*can*”, “*is possible*”, “*is impossible*”, “*should*”, “*should necessarily*” to motivate ourselves. *Modal operators of necessities* are: “*must*”, “*is necessary*”. *Modal operators* refer to how an individual builds his/her own representations in the outside world and define the borders of our *map* and our own style of intervening there.
- *Assumption*—refers only to a moment when the action will take place. This is a linguistic fundament, so that a certain statement makes sense. *assumptions* in a communication process work indirectly and unconsciously and must be accepted in order to give meaning to the communication. These are a form of influence. Used efficiently, they can help achieve results in a short period of time. The value of knowing these patterns consists of that the individual becomes aware of the optimum processes that work within them.
- *Cause-effect relation*—connects the two elements and can be implicit as well as explicit. The two experiences are based on a verbal description. The use of this relation in the communication process helps the weaker person to assume his/her own feelings, states and experiences, thusly facilitating the adoption of an active attitude, by taking into account as many possibilities as possible.
- *Complex equivalency*—assumes the finding of the *breaching word* and reinserting it into another phrase, without changing its meaning. “*The complex equivalency represents a pattern*

with a particular structure, by means of which a person's behavior can be identified based on verbal and non-verbal signals" [6, p. 117]. All signals (verbal and non-verbal language) issued by an individual in a certain period of time, plus the complex equivalency pattern will lead, in the end, to obtaining customized information about that individual by the manager. This information is, generally, distorted by the own feelings as well as by the state of mind at that moment of the respective person. Building a complex equivalency is based on certain words, such as: "is...", "this means...", "is equivalent to...". In the case of a complex equivalency, we create, on mental level, a relationship between one or more words and the experience designated by those words, for each learned word having a different inner experience.

- *Universal quantifiers*—"allow the achievement of generalizations in the communication process, but at the same time, they allow their increasing complexity" [2, p. 97]. The universal quantifiers "always", "everything", "each", "never" represent a form of deleting information. Universal quantifiers have the role of abstracting received information, therefore the generalizations are very hard to achieve.
- *The importance of sensory experience*—orientation of eyes, tone, mimicry, gestures can give the manager and leader enough information to make a decision or they can be used in the case of a too little communicative individual.
- *Exterior behavior*—includes actions and/or events we perceive from the exterior environment. *The significance* (inner state) represents the interpretation we give to a certain event or a certain action in the exterior environment. *The static image* of an exterior behavior means its transformation in semantic construction that action materializes in, leading, in the end, to *substantivization*. The *substantivization* represents the transformation of a static image into words identifying that action, therefore achieving a *cause-effect relation*.
- *Mind reading*—is a model correlated to *assumptions*. The use of *mind reading* assumes the existence of a *cause-effect relation* between two persons, but also of this instrument's effect on the targeted person. In this process, many things can be found about the inner experience of the respective person. The person trying to achieve this process involuntarily projects his/her own perceptions, value, experiences. *Mind reading* appears when an individual speaks or acts as if he/she knows the inner experiences of the other person.
- *Judgments*—in its assessment we must consider the categories used by the person who made the *judgment*. Thus, the questions "Who?", "What?",

"When?", "How?", "Where?" can be used to successfully recover the desired information.

The meta model leads to explaining contexts in which a person manifests himself/herself [12, p. 56], being an efficient instrument not only for gathering information, but also for clarifying their senses and significances. In a purely physical sense, this model becomes a "system of principles, rules and means of knowledge and transformation of reality" [16, p. 137]. The NLP suppositions are associated with assumptions *imprinted* on the structure of an affirmation or subconscious beliefs, actions or convictions, being necessary in order to give sense to them and whose purpose is to allow the person inner development and perception of opportunities and advantages in new, different situations. "Knowing what the meta model means and how to react to them, the meta model can lead to an easier understanding of a person's profound structure" [9, p. 167].

5 METAPHOR

In the book "Encyclopedia of systemic NLP and NLP new coding", Robert B. Dilts and Judith D. Lozier define metaphor as "the transfer of relation between a set of objects and another, with the purpose of obtaining explanations". Also, the metaphor is "the figure of speech in which one speaks about something as if it is something else" [5, p. 226]. The metaphors are memorable and may contain emotions as well as lessons. Humor is ideal for metaphor. A remarkable metaphor can lead the audience through several emotions, from humor to sorrow [13, p. 152].

The metaphor represents a linguistic structure in which one speaks about something as if it is something else [6, p. 87], this "as if" implying two main components, namely:

- *Isomorphism* – the term, used by some linguists to name the structure comparison between syllables, words and sentences, considers the similarities of different representations. *Isomorphisms* have a structure similar to the one of comparisons, using words such as "like", "or", "as". The *isomorphic metaphor* presents analogies with their narration as eloquently and realistic as possible.
- *Symbolism*—considers the structural aspects of metaphor. Thus, the *symbol* may be defined as "an object, a character covering certain behaviors, answers, but generating anchors for others, as he/she/it is part of the inner reality of the concerned persons" [10, p. 53]. The *symbol* reports to any object, situation or person becoming an anchor for certain answers, within a metaphor passing from surface structure language to its deep

structure. also, in a *metaphor* it is not important what object or situation was selected, what is important is how various *symbols* are connected to each other and can interact with the behavior of the concerned individual, having the capacity to facilitate communication. This element has a dominant role in the *metaphor*, so that all relations from that specific *metaphor* and all elements it meets refer to that specific element.

The metaphor is the element used in outlining individual realities. When an individual uses a *metaphor*, he/she gives it his/her own interpretation. The main advantage of a *metaphor* consists of the fact that it can easily overcome the barriers of consciousness, reaching the subconscious level where a useful solution is much easier found for arising problems, solution based on knowledge, skills and experiences belonging to that individual. The subconscious gives certain significance to each thing in the inner or outer environment, but the *metaphor* manages to overcome this obstacle. *“an attribute specific to metaphors is that they facilitate communication. When the metaphor is applied on profound structures of experience, what was transferred or brought by the metaphor are relations, the place of attention, values, beliefs, assumptions”* [10, p. 53].

Metaphors can be: words, expressions, stories. They are based on the idea of comparison, and with the its help, they try to make information as accessible as possible [14, p. 95]. Thusly, by creating comparison, the information from which it starts becomes more accessible to the person who made the *metaphor*.

Metaphors are used for decoding reality by any individual, and are classified into two categories, respectively: *surface* (associated to *isomorphisms*) and *profound* (they have a structure similar to relations, processes, forms, being much more abstract). Through applied directing language models, *metaphors* generate certain differences on language level, with the help of linguistic *markers*, certain old perceptions being able to generate new ones with the role of redirecting main senses. *“The advantages of using metaphors are: they address the unconscious and suggest solutions for solving a certain problem”*. This way, they help individuals identify the defects on behavioral level, they improve elements necessary for the *reframing* process, they reduce the individual's resistance in the communication process. By using *metaphors*, behaviors become more flexible and thusly optimize the communication process [10, p. 55].

6 MILTON MODEL

This model is nothing but the inverted image of a *meta model* as it, in comparison to the *meta*

model, uses an ambiguous, nonspecific language [5, p. 129]. Using language means passing from conscious to unconscious, for the purpose of reducing the rational control of the consciousness, so that the resources states are easier to access and modify in order to obtain the desired result.

Milton model uses hypnotic techniques of Milton H. Erickson (psychiatrist and psychologist), considered the father of modern hypnotherapy, being said about him that he exerts an enormous influence, similar to the one exerted by Sigmund Freud (neuropsychiatrist) in his time. This model offers the user a language structure that is almost completely from contents [11, p. 151]. Using this model allows the access of unconscious resources a person has at a given time and is based on the succession of messages having as purpose the production of a deformed state of the conscious, through hypnosis, thusly reducing the control of the individual's reasoning and his/her capacity to access the resource states, thusly being indirectly guided in the desired direction. The succession of messages assumes the following stages [8, p. 430–451]:

- *fixation of attention* (selective orientation and concentration of psychic activity on certain stimuli or tasks in order to obtain an optimum perception, proper solving of tasks, problematic situations and adaptation of sensory-motor, cognitive and affective behavior to the mobility of outer conditions and the dynamics of person's reasons and purposes);
- *Depotentiation of normal habits* (however, the conscious mind will continue only with the consent of the subconscious);
- *Inserting stimuli of subconscious* (various subliminal messages only the subconscious can recognize);
- *Stimulating positive reactions* (which stimulate reflection and dialogue).

The principles of *Milton model* are [3, p. 129]: *any person has his/her own inner map, is liable for his/her own choices, can identify certain resources to change, must be understood through his/her own visions about the world, any person communicates and the more flexible a person is, the easier he/she can communicate*. The language structure of this model is based more on hypnotic language elements and less on concrete language ones. The hypnotic language is characterized by several specific elements: *it stimulates the conscious as well as the unconscious of a person, it helps overcome conscious barriers and it avoids the appearance of new resistances in the communication process* [11, p. 136]. In order to access a person's hypnosis state, it is necessary to use a certain ensemble of visual, auditory and kinesthetic representations, being mandatory

that all that person's representational systems are involved – simultaneously – in this process.

Milton model is formed of: causal links, ambiguity, fixed orders. a causal link can be seen as a direct report between two events. Ambiguity represents the words, orders with multiple meanings. They can be used for distraction or interruption of the conversation to transmit a message on subconscious level. Phonetic ambiguity targets a certain word that cannot be separated from the rest of the sentence. Syntactic ambiguity targets the partial overlap of two sentences with a common part. By changing the tone, any message can become an order [10, p. 97]. all these stages do nothing but modify the information received within the organization and make very subjective and often wrong interpretations appear, based on the fact that the feelings of a person have been attached to a certain information, this information thusly becoming much too personal for that person to be able to dispose of it and accept it is not correct.

7 3M METHOD

Through the *3M method*, the following modifications can be made:

- *Orienting the meta model so that it responds in real time to the environmental problems* by optimizing communication elements and eliminating the one that can give birth to ambiguous interpretations, fact which can be achieved by replacing words and syntactic structures with a questionable content in usual language. The ambiguous words and syntactic structures have in their contents a large quantity of information, which not only do not clarify certain aspects, but generate new understandings, reason for which it is better to avoid these constructions. In the NLP *meta model* one can design a language model based on priorities so that when a person receives information within the organization, only the words or syntactic structures related to this information are perceived, thusly significantly decreasing the risk of finding and using words and syntactic structures that contain ambiguities. Eliminating or marginalizing certain structures within the communication with a certain ambiguity content can lead to optimizing it. On the other hand, the way of functioning of a language model based on priorities would do nothing but filter the elements that provoke disruptions in the communication process. Using the *meta model* one could optimize the communication process, in the meaning that when a person receives information within the organization and tries to find certain references, this model will help him/her request

only information about that reference, so that the use of words containing a certain degree of ambiguity is eliminated in a significant proportion.

- *Using verbal language that also contains elements for action triggering, so that it stimulates the mental activity to insure a resistance in the message after transmitting it.* Using *motivational triggers* aims at improving results obtained with the help of *Milton model*, this objective being achievable through the joint action of both elements, having a special capacity to overcome the conscious barrier. Therefore, by replacing simple words, action verbs, one can increase the efficiency of the *3M method*, as it offers not only an easier transmission of the message, but also an easier storage than in the case of other methods. Using *motivational triggers* in *Milton model* would have a multiplying effect on the hypnotic effect, as they have a greater capacity to transmit specific information, including action orders for usual words or verbs, thusly increasing the efficiency of the *Milton model* and reducing the time used for transmitting certain orders. Regarding communication elements that compose the structure, they can be optimized, which means that those elements containing ambiguities can be eliminated to improve communication. By using the *3M method*—based on communication priorities—information can be directed to a certain reference, thusly eliminating the possibility of using words with ambiguous meaning. also, for improving the efficiency of using this method, optimized language and the one using elements from Ericksonian hypnotherapy techniques can be combined with the help of *motivational triggers*, the improvement of the quality of perceptions having as purpose the improvement of information collected from surrounding environment and submitted to the filtration process.
- *Improving metaphors*, considering that metaphors have an important role in achieving perception. The improvement of the representation system having as purpose the diversification of perceptions using *metaphors* and the way they are built will pursue the achievement of a good codification of reality; using values, beliefs, customs will aim at increasing relations between values and the identity of the concerned person, which will be much more rooted in the organization through information he/she obtains. The improvement of *metaphors* and how they are built will pursue the achievement of a more performing codification of information within the organization; the more the *metaphors* become efficient, the better is the information codified. Using the widest scale of beliefs, customs, values will be pursued in creating a much

stronger connected between a certain person and the organization.

In the *3M method*, the representation structure consists of identifying the best *anchor*, so that the information within the organization can be associated with a stimulus of *motivational triggers*, according to a certain context or as response to a specific request. Thusly, depending on the *anchors* and their triggering period, information within the organization is prioritized for completion. Also, information already *anchored* is ordered depending on certain criteria set by unspecific indexed references. By inserting certain different *motivational triggers* in the communication process, this entire structure is submitted to an acceleration process. These triggers aim to overcome certain communication routines or overcome certain communication barriers a person may encounter, thusly obtaining efficiency in communication, as well as achieving certain objectives previously set and which must be achieved in the communication process.

8 CONCLUSIONS

The 3M method aims at:

- *Improving the communication process*—leads to improving the organizational performance, improving performance on team level, eliminating semantic construction generating misunderstandings and an easier exchange of messages in an organization or a team;
- *Eliminating as many ambiguous elements as possible in the communication process*—will lead to a faster understanding of the transmitted message and an easier reception of it by the recipients;
- *Directing the communication process to a certain reference, for achieving the objective*—the process of searching unspecific indexed references will lead to their ranking depending on their importance, based on priority, according to their representativeness; thusly, the references must simultaneously meet at least two criteria, respectively to be as relevant as possible and as new as possible, through these references the communication process being improved, and the new information within the organization being updated with the most important references related to a certain aspect of a certain period of time. The orientation to a certain reference will lead to the creation of a stronger connection with the used *anchor* (with the role to access inner features of recourses) as well as to the creation of a stronger connection between a certain reference (image, sound, moment) and the representation system (with the purpose to

extract in certain situations the most complete and relevant information).

Also, by improving *metaphors*, the way they must be built, the used analogies will lead to improved solutions for real problems and which a person might find on subconscious level, and the increase of perceptions will lead to creating closer bonds with the representation system and obtaining information regarding a quality of the employees' professional life. Thusly, the communication process is significantly improved, optimized, the use of ambiguous elements is reduced, references are classified depending on their importance and priority, stronger connections are made between the representation system and references and the uses of analogies are pursued, so that the solutions to various problems have, in most cases, an inner source, not a source outside the organization.

REFERENCES

- [1] Andreas S. & Faulkner C. (2008), "*NLP și succesul*" ("*NLP and success*"), Curtea Veche Publishing, Bucharest;
- [2] Bandler R.W. (1993), "*Time for a change*", Capitola: Meta Publications Inc.;
- [3] Bandler R.W. & Grinder J.T. (1982), "*The Structure of Magic. Volume I*", Palo alto: Science and Behaviour Books Inc.;
- [4] Bandler R.W. & Grinder J.T. (1982), "*Reframing*", Moab: Real people Press;
- [5] Dilts R.B. (2007), "*Bazele programării neuro-lingvistice*" ("*Basis of Neuro-Linguistic Programming*"), Excalibur Publishing, Bucharest, 2007;
- [6] Dilts R.B. (2008), "*Strategii de geniu, Volumul I*" ("*Strategies of genius. Volume One*"), Excalibur Publishing, Bucharest;
- [7] Dilts R.B. & Lozier J.D. (2000), "*Encyclopedia of Systemic NLP and NLP New Coding*", NLP University Press: Scotts Valley;
- [8] Erickson M.H. & Rossi E.L. (1980), "*Two-level Communication and the Microdynamics of Trance and Suggestion*" in E. L. Rossi (Publishing)—"*The Collected Papers of Milton H. Erickson on Hypnosis: Vol. 1. The Nature of Hypnosis and Suggestion*", New York: Irvington, p. 430–451;
- [9] Hall L.M. (2007), "*Spiritul programării neuro-lingvistice*" ("*Spirit of Neuro-Linguistic Programming*"), Curtea Veche Publishing, Bucharest;
- [10] Iosif C.M. (2013), "*Utilizarea performantă a programării neuro-lingvistice (NLP) în managementul firmei*" ("*Performing Use of Neuro-Linguistic Programming (NLP) in Company Management*"), C. H. Beck Publishing, Bucharest;
- [11] Knight S. (2007) "*Tehnicile programării neuro-lingvistice*" ("*Techniques of Neuro-Linguistic Programming*"), Curtea Veche Publishing, Bucharest;
- [12] Lewis B., Pucelick F. (1980), "*Magic NLP demystified*", Portland: Metamorphous Press, 1990;

- [13] Molden D. *“Business Masterclass: Driving Peak Performance with NLP”*, London: FT Press, 2007;
- [14] O’Connor J. (2012) *“Manual de programare neuro-lingvistică. ghid practic pentru obținerea rezultatelor pe care le dorești”* (*“Neuro-Linguistic Programming Manual. Practical Guide for achieving the Results You Desire”*), Curtea Veche Publishing, Bucharest;
- [15] Vagu P. & Stegăroiu I. (2007), *“Motivarea în muncă. De la teorie la practică”* (*“Motivation in Work. From Theory to Practice”*), Bibliotheca Publishing, Târgoviște;
- [16] Zaiț D. & Spalanzani A. (2006), *“Cercetarea în economie și management. Repere epistemologice și metodologice”* (*“Research in Economics and Management. Epistemological and Methodological Views”*), Economica Publishing, Bucharest.

Business Process modeling: Case of undergraduate program

K.V. Zhukova

Information Technologies in Management Department, Graduate School of Management Saint Petersburg State University, Saint Petersburg, Russia

A.Yu. Pleshkova

Graduate School of Management Saint Petersburg State University, Saint Petersburg, Russia

ABSTRACT: Numerous business actions in university practice may lead to confusion and can significantly reduce the effectiveness of the administrative and educational processes. Hence, this paper aims to present how Graduate School of Management of Saint Petersburg State University (GSOM SPbU) enhanced its Business Processes (BP) and Knowledge Management (KM) by developing the model of information support for the undergraduate program Directorate and undergraduate students. The study presents the results of applied project that included the model development for information support of the activities of undergraduate programs Directorate, a visual representation of the scheme of business processes, directory structure for storing documents and students' academic calendar. Based on the in-depth interviews paired with content analysis the study describes the framework of business processes improvement. The project helped to create the environment that efficiently employs business processes thus enhancing the efficiency of educational community. Other universities or educational institutions can use this BP and KM experience of GSOM SPbU as an approach to modeling the administrative and educational processes.

1 INTRODUCTION

The ongoing advance of technology has influenced in people's actions and activities to become more complicated with likelihood of appearance of the new problematic situations. The educational sphere as well as precisely the sphere of high schools is not an exception.

Improving process management becomes a key objective of many companies. Low level of strategy success is primarily due to the fact that information maintenance and analytical support did not pass into the category of more or less proven technologies (Latunin & Bokova, 2003).

The problem of information support is getting into two science spheres and these are Knowledge Management (KM) and Business Process Management (BPM). KM allows companies and institutions to manage and exchange knowledge from the place where it is originally generated to where it is to be exploited. KM assists the needs of internal (in our case, the faculty, teachers, administrative staff, etc.) and external customers (in this case, students) by generation of organizational routines that facilitate creativity of individuals and effective processes (Marulanda & Montoya, 2015).

Papers regarding the issues of business processes and educational context tend to differ in the ways they are trying to look at the problem in order to

solve it. Some (Gjoni, 2015) use the approach to develop information systems and implement a model-oriented approach since it focusses on the business logic (the "what") rather than on the specific implementation technology (the "how") (Lacerda et al. 2014).

BPM is not only about designing, developing and executing business processes, it also considers the interaction between these processes, managing, analyzing and optimizing them (Kohlbacher, 2010; Saraswat et al. 2014). Changing the approach to operational management of the company to a process oriented management approach involves defining the responsibilities for the conduct of the proceedings (Ahmad et al. 2007; Palmberg, 2010), minimize transfers, thereby reducing errors and time delays, maximize the grouping of activities and reduce the effort (Antonucci & Goeke, 2011; Paim et al. 2008).

Organizational design of business processes is a leadership competency and responsibility that is taking on even greater importance as organizations require agility to respond to the environment (Ritacco, 2015). Human resources with developed talents and creativity who are able to reach and utilize information constitute the main power of competition in the world market (Kleinhempel et al. 2010; Cabanillas, 2016). Those companies and institutions that make investments on human resources

and attempt to create working conditions that are compatible with their requirements and wishes, are the ones who reach success (Burma, 2014).

These changes are of top importance if institution desires to remain competitive. As means of support companies can employ various approaches, techniques, tools and models; these, however, are not always adapted to the needs (Vedenick & Leber, 2015). In our case we have identified them before constructing the scheme of business processes and their implementation.

Business processes in the context of the current available Information Technologies (IT) leads business education towards sustainable development and highlights its ability to offer a missing link between business, IT and strategy (Seethamraju, 2012). According McCormack et al. (2009), advancing in the management of business processes, the organization will have better control of results, better prediction of goals, cost and performance.

There are permanent requirements for the changes in performances, increasing flexibility and improving the economic position of the company or other institution through the process orientation (Milan et al. 2014). As processes are aimed to the same goal, unnecessary and misdirected steps are redesigned or eliminated, concentrating resources on core processes and improving the organization's performance (Segatto et al. 2013) and the systemic approach may be a key subject to clarify the inter-relationships among processes (Basal, 2010), and processes and their contexts. In this paper we have considered different approaches towards constructing the administrative and academic business processes.

Some research specify on the focused problems of curriculum upgrades in one educational program (Hauck, 1998) but in our paper we describe the case of dealing with 18 business processes regarding undergraduate program from two different perspectives.

Other research findings upon the curriculum design (Lin, 2015) present partially positive effects on fluency, flexibility, originality, and elaboration and reveal significant moderating effects on the correlations between curriculum design and creative potential developing (Vazzana et al. 2000). Effective business process management inside the institution in its turn allows to enhance the overall quality of the knowledge management policies (Cao, et al. 2013).

The managerial problem in GSOM SPbU that we analyze is typical for all educational institutions. Due to the changes in the organizational structure on the level of SPbU, insufficient staff and changes in the undergraduate programs Directorate, the integrity of the information had been flawed and

that ultimately led to the need of developing a model of information support. This paper will describe how information support was maintained inside GSOM SPbU regarding the management of the undergraduate programs.

The remainder of this paper is structured in the following way: we analyze the circumstances that led to the emergence of the problem, develop the framework to work with the problem, provide the example of the particular case and summarize main outputs in the conclusion part.

2 EDUCATIONAL TRENDS IN RUSSIA

2.1 *SPbU organizational changes*

After September 2009 SPbU initiated the process of the departments transformation—all departments of SPbU started to unite into separate branches and GSOM SPbU became a part of the branch “Geology and Management”. Each branch now was subordinating to the Vice rector of Academic affairs. These organizational changes were completed to increase the efficient use of the resources and reduce costs for individual control of every department.

2.2 *GSOM SPbU organizational changes*

Graduate School of Management of Saint Petersburg State University (GSOM SPbU) is a recognized leader in Russian business education. GSOM SPbU is one of the Departments of SPbU and is submitting to SPbU. Up to the 2009 GSOM SPbU had divisional organizational structure in which the departments of the University were the divisions. Directorate of the undergraduate programs was submitting to the Dean, Office of the undergraduate programs was submitting its Director.

In 2014 the basic principle of GSOM SPbU organizational structure was separation by the type of activities. This organizational structure can be called “matrix”.

GSOM SPbU has been ranked by EDUNIVERSAL as #1 business school in Russia for 6 consecutive years since 2008 and #1 business School in Eastern Europe since 2012.

After 20 years of dynamic growth GSOM SPbU has gained an unprecedented for a Russian business school international recognition through a set of institutional memberships in the most prestigious international professional associations. The School today is the only Russian business school to be accredited by both AMBA and EQUIS (EFMD), the only Russian member in the Global Alliance in management Education (CEMS) and Partnership in International Management (PIM).

International reputation of GSOM SPbU is also confirmed by a unique and strong network of international academic partners. 60 partner business schools from Europe, Asia, Australia and the Americas are among top-3 business schools in their respective countries. This all means high level of international responsibility and expectations towards the quality of the service.

We can outline six main circumstances on the level of GSOM SPbU that led to the emerging problem of the need for new information support model and these are:

1. Implementation of the normative and regulative acts variation in departments of SPbU;
2. Era of big change at SPbU and a strong need to respond quickly and to build business processes tailored to the new organizational structure;
3. Need to support the particular quality level of business processes due to the requirements of international accreditations (see Appendix).
4. Shift from the department towards the program management principle;
5. Increase of the complexity of organization of educational processes under the certain conditions of world accreditations (students exchange program, internship exchange program, etc.);
6. Complexity of the planning and interconnections between the processes under the conditions of high document volume.

They are typical for the changing nature of the educational sphere and are likely to happen in different educational institutions. These fundamental changes led to the initiative of the new business process model development at GSOM SPbU.

3 NEW MODEL DEVELOPMENT

3.1 *Challenges to consider*

The problem of informational support for undergraduate programs splits into two perspectives: Directorate and students parts. This splitting allowed to work upon one managerial problem but from two different perspectives. For both of them are three major needs to be fulfilled:

1. Creation of information space for support of business processes;
2. Development of tools for planning Directorate activities;
3. Development of academic calendar for planning student activities.

3.2 *Methodology*

The applied project employed qualitative method that consisted of in-depth interviews with key

managers of GSOM SPbU and content analysis of the regulatory documents of GSOM SPbU (Pehlin, 2014; Mihnevich, 2014). This allowed to gain valuable administrative insights, form and refresh the database of normative documents and get the understanding of the processes flow.

3.3 *Approaches towards modelling business processes*

During the process of the plan development numerous ways to allocate resources upon certain time limits appear. To make this process clear and understandable for everyone we should base on the four key selected elements and in our case these are: events, documents, participants of the process and time scale.

We can choose between four main approaches regarding the construction of the overall big plan and these are: process approach, HR approach, complex approach and improved complex approach (used in our case). While choosing between these approaches we have to understand the goals of the final result and consider the limitations of each approach. The main limitation of the process approach is hard perception. We have the processes on the vertical axis and time on the horizontal one. But because some people are involved in different processes simultaneously there will be a large number of duplicate rows for the participants of business processes. Moreover, the line will appear to be too busy because of a large number of documents in a short period time.

HR approach differs from process in the way of presenting the information: the pushing off point (vertical axis) is for people while the horizontal remains the same. But major drawback of is the confusion in detecting the needed documents. First, the absence of the document-line does not allow directly (without passing through a hyperlink) to see which documents are involved in the process. Secondly, when a large number of processes are presented the picture will become complicated. Additional problems may arise if one participant will have different actions on multiple processes in a single day.

In the complex approach the horizontal axis (time) is divided into the educational weeks and enriched with the important dates (such as dates of the department meetings, commissions, etc.). Vertical axis also has changes—in this approach it outlines the descriptions of the ongoing actions. The disadvantage of that approach is inability to describe exactly what issues were discussed at a particular commission. Analysis of the possible approaches led to the choice of the fourth—improved complex approach. The horizontal axis stays for the working weeks and the vertical is for

the business processes. By adding hyperlinks to the needed documents and storing them in particular folders in database we can minimize the time costs and increase efficiency.

4 BUILDING THE BUSINESS PROCESSES

To start with we have defined 18 key business processes. A large number of participants in the proceedings of business processes greatly complicates the implementation and execution control. This determined the need for beforehand planning of the operations of undergraduate programs Directorate, both in relation to employees and teachers and to the students of undergraduate programs.

Therefore, we have one managerial problem to solve regarding two perspectives: Directorate of the undergraduate programs and student's perspectives. From the students perspective we developed a model of information support for 8 major business processes (preparation/organization/fulfillment) and also the academic student calendar:

1. Questionnaires about the quality of teaching;
2. Profile distribution;
3. Start of the semester in "Blackboard" SPbU system;
4. Organization of the State Attestation Commission;
5. Registration for the choice disciplines;
6. Educational process;
7. Diploma preparation;
8. Graduation ceremony.

From the undergraduate programs Directorate perspective managerial problem was to develop the model regarding other 10 key business processes. List of business processes can vary in different high education schools but usually it presupposes the following positions (preparation/organization/fulfillment):

9. Students exchange;
10. Internship exchange;
11. Draft standards of SPbU;
12. Study plans;
13. Teaching assignments volume;
14. Syllabus development
15. Reinstatement and transfer;
16. Bachelor term-papers and thesis development;
17. Invited professors recruitment;
18. Freshman day.

Then the initial challenge was to develop the administrative schedule that will improve control over the order and timing of the execution of business processes throughout the school year and capture the big picture. Information support of the operations of management of the undergraduate programs basically consists of three major

elements: Storage of documents, schemes of visual representation of business processes (in our case developed on the basis of Microsoft Excel) and the administrative timetable of the directorate of undergraduate programs. To achieve the goal there is the need to solve the following sub challenges:

1. Analyze the existing tools to build the model of information support;
2. Develop a visual representation of the scheme of business processes;
3. Develop a framework of the document repository;
4. Plan business processes for directorate of undergraduate programs;
5. Identify peak periods;
6. Develop the administrative schedule for directorate of educational programs.

5 CASE DESCRIPTION

We will focus on one particular example of the information support for one business process of the undergraduate programs Directorate—"Students exchange" business process.

GSOM SPbU has 60 international university agreements. In charge of this business process are employees from Procurement sector and International office. Competition for the exchange semester abroad is held twice a year. Competition is fulfilled through the following algorithm:

1. Order distribution—announcement of the beginning of competition and its key important dates are sent via email to the undergraduates including online publishing;
2. Application acceptance—International office collects the applications filled in with respect to all GSOM SPbU standards (1 week);
3. Application processing—International office processes all the applications:
 - a. Creating a database of students regarding their preferences and scores;
 - b. Meeting of committee. (1 week)
4. Project of the order approving the results of the competition;
5. Nomination of the students and fill-in the online forms of the partner business school;
6. International office gets the invitation from the partner business school and transfers it to the student;
7. Before departure students approve plan for education;
8. Order for study abroad departure;
9. All orders approval;
10. If there are any changes in the study plan student is responsible to send the new one signed by the partner business school;

11. After arrival back to GSOM SPbU student brings copies of passport, visa, etc.;
12. Documents of return:
 - a. Official note if student arrived on time;
 - b. Official note if student arrived long after scheduled date;
 - c. Order if student arrived before scheduled date.

By analyzing these steps we can organize them into one business process and structure all the needed documents into the document repository. It is a set of structured files stored on the server. Access is either by hyperlink in the overall plan in MS Excel or direct. The defining point is the structure of the data. It does not only easily find the documentation for business processes and is convenient for immediate access but also allows to understand all of the steps of the business process at once. Each process has the abbreviation before the naming. For “Students exchange” business process we will have the folder named “SE_Students exchange”. Next the structure of the process is logically subdivided into additional levels (for example, “Planning” folder contains of all necessary documents for the start of the process, there is also the separation for fall and spring semesters, the last document catalogue in this case will be “Results”).

6 CONCLUSION

In this article we described how to unite the two perspectives of one common managerial problem and solve it through applying the effective use of the BPM and KM policies. The major output is the developed sequence of actions that high school or other educational institution has to proceed:

1. Problem identification
 - a. Identification of key participants
 - b. Identification of the expectations of key participants
2. Preparation towards problem solution
 - a. Identification of business processes
 - b. Identification of their participants and responsible departments
 - c. Set of time limits
 - d. Collection of all normative and regulative documents
3. Problem solution
 - a. Building the database of normative and regulative documents
 - b. Constructing the structure of access to them
 - c. Planning the administrative calendar (getting the big picture)
4. Initiation.

The main outputs of model development of information support for student processes were:

- List of business processes;
- Development of forms to describe business processes in the Excel spreadsheet format;
- Development of structure of information space on the basis of the file system directory;
- Description of the business processes on the basis of the submitted documents;
- Conducted analysis of business processes;
- Development of student’s academic calendar.

The major outputs of model development of information support for Directorate processes were:

1. Analysis of the available tools for building the model of information support;
2. Development of the visual representation of the business processes scheme;
3. Plan of the business processes of Directorate of the undergraduate programs;
4. Conducted analysis of peak loads in business processes;
5. Development of Directorate’s administrative calendar.

We had the experience to picture all main business processes for undergraduate programs in GSOM SPbU. This generalized representation served to determine the period of greatest workloads regarding the management of undergraduate programs (see Figure 3). This is a crucial moment in the overall planning process and it allows to understand where intensive resource allocation is needed.

On the basis of planned business processes administrative schedule was developed for undergraduate management programs to provide information that supports business processes flowing.

6.1 *Practical implications*

Before the model construction we formulated the primary requirements that can be seen as further benefits of this approach:

- Planning—due to the large number of legal documents with different expiration dates the model of business processes must be constructed with respect to time limits and should comprise top point events in these business processes;
- Need to describe input and output documents in relation to business process—the description of business processes is subject of educational institutions and it is obvious that they will be governed by a large number of input and output documents. In GSOM SPbU there are three document flows: the flow of incoming and initiating documents (“what”), documents regulating the operation of the business processes (“how”) and results documents (“what happened”).

- Possibility of scalability—it can be called by another term—flexibility. The model should be designed for next academic year use. Hence it should be adapted to changes in the regulatory framework and standards.
- Multi-user interface—flexibility can also be expressed in the ability of several people simultaneously to work with the system. This requirement is important because the administrative department of GSOM SPbU consists of a few dozen employees, each of whom is a participant of a particular business process.
- Low cost—the final requirement is that the cost of implementation of this model.

These requirements set the conditions for the model and can be implemented in different educational institutions.

Implementation of the basic business processes varies from year to year, as members of the business processes of the two main educational undergraduate programs (in the direction of “Management” and “State and municipal management”) are 750 students and 80 teachers and 50 GSOM SPbU actively involved in different parts of the educational process.

The applied project that this paper describes did not aim to detail every action of the Directorate during the 2013–2014 academic year, on the contrary, on the basis of knowledge about the processes and official documents there was the goal to develop a model that describes the information support of the educational process. The goal was successfully achieved and that allowed forming a general idea of the list of objectives to be carried out beforehand in order to maintain the quality of educational services provided to the student.

Considering the practical implementations, we have:

- Ability to observe the links between the organization activities and how to monitor ongoing implementation of business processes;
- The possibility of the analysis of processes and sub-processes and their optimization;
- Clear view on the role of each business process in the whole functioning of the institution;
- Transparency and visibility of the Department’s actions;
- Introduction of the improved process approach towards educational management will improve the quality of the educational services.

To analyze the efficiency of the processes inside educational institute we advise to use the following criteria:

- Time for processes execution;
- Quality of the execution (no delays);
- Costs on resources for each particular process.

6.2 Limitations and further research directions

There are several points on which the researchers can explore more and that were not covered by this applied project:

- Depth of study—although we managed to build the big picture of the ongoing business processes and make the research of high latitude we have not considered the sub processes of each action/operation;
- Influential factors—we have not considered the possible influence of psychological/motivational factors to the fulfillment of the described business processes;
- Focus of research—we analyzed business processes of undergraduate programs Directorate and students and we see potential in research regarding other educational programs (for example, master/doctoral/executive);
- Applicability—we provided detailed description of the construction and implementation of business processes on the particular example of GSOM SPbU but nevertheless there are practical implications that can be used in other educational institutions.

REFERENCES

- Ahmad, H., Francis, A. & Zairi, M. 2007. Business process reengineering: critical success factors in higher education. *Business Process Management Journal*, 13(3): 451–469.
- Antonucci, Y. & Goetze, R.J. 2011. Identification of appropriate responsibilities and positions for business process management success. *Business Process Management Journal*, 17(1): 127–46.
- Basal, A. 2010. Business Process Reengineering towards an Integrated Learning Management System, *Proquest LLC*.
- Burma, Z. 2014. Human resource management and its importance for today’s organizations. *International Journal of Education and Social Science (IJESS)*, 1(2): 85–94.
- Cabanillas, C. 2016. Enhancing the management of resource-aware business processes. *AI Communications*, 29(1): 237–238.
- Cao, Q., Thompson, M.A. & Triche, J. 2013. Investigating the role of business processes and knowledge management systems on performance: A multi-case study approach. *International Journal of Production Research*, 51(18): 5565–5575.
- Conger, S. 2011. Process Mapping and Management. *Business Expert Press*.
- Gjoni, O. 2015. Comparison of two model driven architecture approaches for automating business processes, MOSKitt framework and bizagi process management suite. *Mediterranean Journal of Social Sciences*, 6(2): 615–625.
- Hauck, A. 1998. Construction management curriculum reform and integration with a broader discipline:

- A case study. *Journal of Construction Education*, 3(2): 118–130.
- Holt J. 2009. Pragmatic Guide to Business Process Modeling. *BCS (2nd ed.)*.
- Kleinhempel, S., Niçhi, Ş. & Rusu, L. 2010. Business Process Management in Service-Oriented Companies. *InformaticaEconomica*, 14(3): 189–198.
- Kohlbacher, M. 2010. The effects of process orientation: a literature review. *Business Process Management Journal*, 16(1): 135–52.
- Lacerda, R., Ensslin, L., Ensslin, S. & Dutra, A. 2014. A Constructivist Approach to Manage Business Process as a Dynamic Capability. *Knowledge & Process Management*, 21(1): 54–66.
- Latunin A. & Bokova A. 2003. Model of informational and analytical support for strategic management. *Rossiiskoeprindmatelstvo*, 2(38): 85–92.
- Lin, P.-C. & Lin, P.-K. 2015. Effects of curriculum design on students' creative potential developing—a case study on students in the department of business management. *ActaOeconomica*, 65: 267–277.
- Marulanda, N. & Montoya, I. 2015. Knowledge management and technological innovation capabilities as tools for business performance evaluation. *Turkish Online Journal of Educational Technology*, 362–374.
- Mihnevich, A. 2014. Model development for information scaffolding for bachelor program directorate activities. *Graduate School of Management SPbU*.
- Milan, R., Milan, B., Marko, C., Jovanovic, V., Dalibor, B., Bojic, Z. & Avramovic, N. 2014. Implementation of Business Process Reengineering in Human Resource Management. *Engineering Economics*, 25(2): 211–222.
- Paim, R.C.S., Caulliriaux, H. & Cardoso, R. (2008). Process management tasks: a conceptual and practical views. *Business Process Management Journal*, 14(5): 694–723.
- Palmberg, K. (2010). Experiences of implementing process management: a multiple-case study. *Business Process Management Journal*, 16(1): 93–113.
- Pehtin, I. 2014. Model development for information scaffolding for bachelor program students activities. *Graduate School of Management SPbU*.
- Ritacco J. (2015). Organizational design assessment: a practical tool for creating organizational agility. *International Journal of Education and Social Science (IJESS)*, 2(6): 85–88.
- Saraswat, S., Anderson, D. & Chircu, A. 2014. Teaching Business Process Management with Simulation in Graduate Business Programs: An Integrative Approach. *Journal of Information Systems Education*, 25(3): 221–232.
- Seethamraju, R. (2012). Business process management: a missing link in business education. *Business Process Management Journal*, 18(3): 532–547.
- Segatto, M., de Pádua, S.D. & Martinelli, D.P. (2013). Business process management: a systemic approach? *Business Process Management Journal*, 19(4): 698–714.
- Silva, L. d., Damian, I. M. & !, S. I. (2012). Process management tasks and barriers: functional to processes approach. *Business Process Management Journal*, 18(5): 762–776.
- Vazzana, G., Elfrink, J. & Bachmann, D. 2000. A Longitudinal Study of Total Quality Management Processes in Business Colleges. *Journal of Education for Business*, 76(2): 67–69.
- Vedenik, G. & Leber, M. 2015. Change management with the aid of a generic model for restructuring business processes. *International Journal of Simulation Modelling (IJSIMM)*, 14(4): 584–595.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

A full duplex media access protocol for hybrid visible light communication networks

Xuehui Wang, Lei Zhang & Wenhua Dou

School of Computer, National University of Defense Technology, Changsha, P.R. China

ABSTRACT: Visible Light Communication (VLC) has attracted much attention because of its superior characteristics, including unlicensed wide bandwidth, high security and dual-use nature. Nevertheless, visible light is unsuitable for uplink data transmission, because constantly turning on the visible light can cause visual disturbance to users. To support bidirectional communication for VLC networks, people proposed hybrid VLC network, where downlink channel use visible light while uplink channel use infrared light or wifi, this result in a full duplex channel. In this paper, we propose a MAC protocol FD-CSMA/CA that supports full duplex communication in hybrid VLC networks. In FD-CSMA/CA protocol if an acknowledge frame need to be returned immediately on the same channel during the transmission of a data frame, it will split the data frame, send the control packet first and resume the transmission of the remaining data frame automatically. At the receiver, the split sub-frames are merged together to recreate a complete frame. By this way, both the uplink and downlink can transmit data or acknowledge frames concurrently. We build a hybrid VLC network testbed and implement FD-CSMA/CA protocol in the linux kernel network stack. Performance evaluation shows the average throughput of FD-CSMA/CA protocol is improved about 75% in a three nodes network compared to the CSMA/CA protocol.

1 INTRODUCTION

Visible Light Communication (VLC) utilizes visible light from Light Emitting Diodes (LEDs) to convey digital information between devices. Compared to the traditional RF based wireless communications, VLC possesses a number of appealing advantages. First, the spectrum of VLC is about 400 THz and license free, its potential bandwidth is much higher than the wifi. Second, VLC can reuse existing lighting infrastructure for communication, reduce the deploying cost of the network. Third, due to its high frequency, visible light cannot penetrate through walls, we can create smaller cells and get higher capacity. Furthermore, VLC has no electromagnetic radiations, which provides better security and is suitable for special areas, such as military or airplane communications.

However, one of the problems of VLC is the complicity in upstream transmission, usually visible light cannot be used for uplink communication, because constantly turning on the visible light can cause visual disturbance to users while using the devices. To address this problem, hybrid VLC network is proposed, where Infrared light or wifi are used as upstream sources, especially infrared is an attractive alternative considering its low cost and no harmful to eyes.

A typical hybrid VLC network is depicted in Figure 1, the ceiling lamp is the access point of

the network, it can transmit data to the terminals by downlink visible light channel, each terminal is equipped with an infrared LED, which FOV is directed to the AP to transmit uplink traffic. For this hybrid VLC-IR network, the uplink and downlink channel work at different frequency, they can transmit signal simultaneously without interference. Since the uplink channel is shared by all terminals, a media access protocol is needed to avoid collisions.

There are some challenges to design a MAC protocol for this hybrid VLC-IR network:

1. Traditional CSMA protocol only support half duplex communication, which is suitable for single channel wireless networks, such as wifi, the wireless channel is shared for the downlink and uplink traffic. But for the hybrid VLC network, the downlink and uplink can transmit and receive at the same time, traditional CSMA based MAC protocol cannot utilize the full duplex channel bandwidth of hybrid VLC network.
2. Because of the FOV directionality of “infrared antenna”, VLC terminals usually cannot communicate with each other directly. In fact in hybrid VLC network they even cannot hear each other because the receiver and transmitter working at different frequency, therefore collision avoidance mechanism is important for uplink channel.

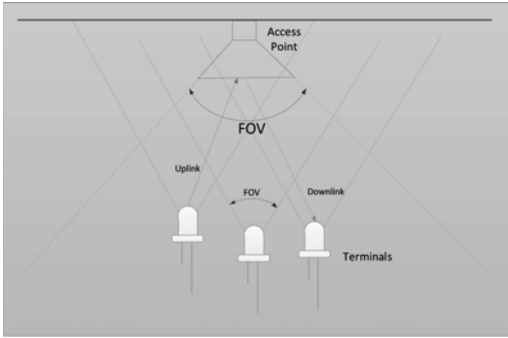


Figure 1. A typical hybrid visible light and infrared communication network.

3. Link layer acknowledgement is necessary to support reliable communication for VLC networks, because visible light communication is very sensitive to receiver movement and orientation, the drop in received optical power can be significant due to such misalignment, which lead high packet lost rate in practical environment, link layer acknowledgement is helpful to improve network throughput.

To meet these challenges, we propose a new full duplex MAC protocol FD-CSMA/CA for hybrid VLC networks. In FD-CSMA/CA protocol during the transmission of a data frame, if there is a control frame need to be sent immediately (such CTS/ACK) on the same channel, the data frame will split into two frames, sending the control packet, and resume the transmission of remaining data frame. At the receiver, the split sub-frames are merged together to recreate a complete frame. By this way, both the uplink and downlink can transmit data simultaneously, sometimes data is interrupted by short control packet and resume automatically. We build a hybrid VLC testbed and implement FD-CSMA/CA protocol on linux platform, Performance evaluation shows the average throughput of FD-CSMA/CA protocol is improved about 75% in a three nodes network compared to the CSMA/CA protocol.

2 RELATED WORK

There are two types of random channel access mechanisms proposed by IEEE 802.15.7 standard [1]. The first type is an unslotted random channel access with CSMA. The second type is a beacon enabled protocol and the time is divided into beacon intervals. A superframe within the beacon interval contains Contention Access Periods (CAP) and Contention Free Periods (CFP). IEEE 802.15.7 is design for single channel visible light networks.

[12] proposed a CSMA based MAC protocol for LED-to-LED communication network, it assumes each node can receive others light, which is the basis for carrier sense, but in most realistic scenarios, the FOV of each node cannot cover all the other nodes.

[14] proposed a Carrier Sensing Multiple Access/ Collision Detection & Hidden Avoidance (CSMA/CD-HA) Medium Access Control protocol, it introduces the intra-frame bidirectional transmission with only one Light Emitting Diode (LED) to transmit and receive data, the embedded reverse communication channel not only improves the network throughput but also alleviates the hidden terminal problem, but it depends on the modulation scheme in which there are data symbols without emission of light.

Although some authors have proposed the use of visible light and infrared light or LEDs operating on different wavelengths for “full-duplex” communications[3][9][20], to the best of our knowledge, there is yet no work on full-duplex MAC protocols for hybrid visible light communication networks.

3 PROTOCOL DESIGN

We first briefly introduce the background information employed in this work, and then present our method to enable full duplex communication for hybrid VLC Networks.

Traditional Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) protocol is depicted in Figure 2. When the AP or terminals wish to transmit, they first waits for a random back-off period and then senses the channel to be busy or not before transmitting. If the channel is found to be busy, they wait for another random period before trying to access the channel again. If data is successfully received by destination, it will return an acknowledgement to the source. If there are collisions and the data is lost, it will use a random exponential binary back-off algorithm to retransmit the data. To alleviate the hidden node problem, a RTS/CTS handshake mechanism can be employed.

3.1 FD-CSMA/CA protocol

In CSMA/CA MAC protocol, at any time only one channel direction can transmit packet, otherwise

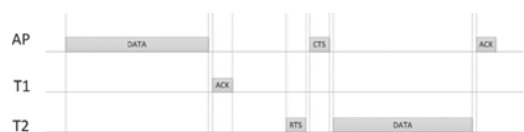


Figure 2. CSMA/CA protocol.

the CTS/ACK packet may collide with the data, which means it only supports half duplex communication. For hybrid VLC networks the uplink and downlink channel working on different frequency, the full-duplex channel bandwidth is wasted.

To allow simultaneous data transmission on uplink and downlink, we must deal with the collision between DATA and ACK/CTS packet. Our basic idea of FD-CSMA/CA protocol is data frames can be dynamically split into multiple sub-frames during transmission, and control packet can be inserted between two split data frames.

First consider the downlink transmission is split by the uplink transmission, as depicted in Figure 3. Suppose AP is transmitting DATA to T1, now T2 has data to transmit to AP, it senses the uplink channel and find it's clear, then it sends RTS to reserve the uplink channel. The AP receives the RTS and it need to return a CTS immediately, it first stops the data transmission, record the data length it has transmitted, send back the CTS to T2, and continue transmitting the remaining data. When T2 receives the CTS, it begins to send data to AP, now both the uplink and downlink is transmitting data. When T2 finishes its data transmission, it waits for an acknowledgement from AP. If AP receives the data correctly, it will stop the data transmission and send an ACK to T2, then it resumes the data transmission again. When T1 successfully receives the data from AP, it will return an ACK to AP, no collisions occur during the bidirectional data transmissions.

The uplink data can be split by downlink data similarly, as depicted in Figure 4. T2 is transmitting data to AP and AP is transmitting data to T1, AP finishes data transmission earlier and waits for an ACK. Meanwhile both T1 and T2 receives the data, T2 knows another terminal need send an ACK to AP immediately, so he stops transmission data and releases the uplink channel temporarily, after T1 returns an ACK to AP, it will continue transmitting the remaining data.

First consider the downlink transmission is split by the uplink transmission, as depicted in

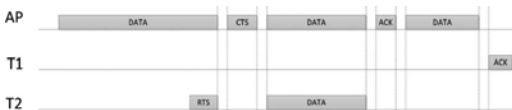


Figure 3. FD-CSMA/CA: Downlink is split by uplink.

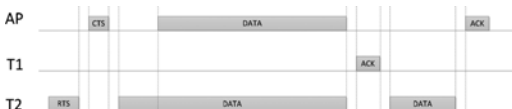


Figure 4. FD-CSMA/CA: Uplink is split by downlink.

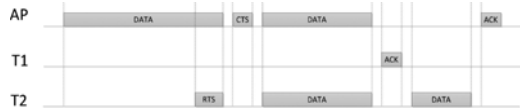


Figure 5. FD-CSMA/CA: A more general split case.

Figure 3. Suppose AP is transmitting DATA to T1, now T2 has data to transmit to AP, it senses the uplink channel and find it's clear, then it sends RTS to reserve the uplink channel. The AP receives the RTS and it need to return a CTS immediately, it first stops the data transmission, record the data length it has transmitted, send back the CTS to T2, and continue transmitting the remaining data. When T2 receives the CTS, it begins to send data to AP, now both the uplink and downlink is transmitting data. When T2 finishes its data transmission, it waits for an acknowledgement from AP. If AP receives the data correctly, it will stop the data transmission and send an ACK to T2, then it resumes the data transmission again. When T1 successfully receives the data from AP, it will return an ACK to AP, no collisions occur during the bidirectional data transmissions.

The uplink data can be split by downlink data similarly, as depicted in Figure 4. T2 is transmitting data to AP and AP is transmitting data to T1, AP finishes data transmission earlier and waits for an ACK. Meanwhile both T1 and T2 receives the data, T2 knows another terminal need send an ACK to AP immediately, so he stops transmission data and releases the uplink channel temporarily, after T1 returns an ACK to AP, it will continue transmitting the remaining data.

A more general scenario is depicted in Figure 5, where downlink data is interrupted by CTS and then uplink data is interrupted by ACK.

3.2 Data frame split and merge

The key challenge for FD-CSMA/CA protocol is how to dynamically split the data into sub-frames at the transmitter and merge them correctly at the receiver. It is relatively simpler for transmitter to split frames. Once there is a control frame (CTS or ACK) to send, the transmitter can stop the data transmission immediately, record how many data bits has been successfully transmitted, wait for a guard interval and begin to transmit the control frame. After the control frame is transmitted, insert a guard interval again and continue transmitting the remaining data.

Because the data frame can be split at any position, the transmitter do not know when the frame will be split, so the frame length field in frame header can only fill the remaining data length, not the current sub frame length. Usually the receiver depends

on the frame length field to determine how much data to receive, this does not work for the split sub-frames. To address the problem, we utilize the physical carrier sense mechanism and remaining data length together to correctly receive and merge the data sub-frames. There is a guard interval between two consecutive frames, so the receiver will detect the channel is clear during the guard interval, which can be used as a sign of sub-frame end to stop receiving. One problem is carrier sensing is not so precise, usually there is a delay from the guard interval beginning to carrier sensing channel clear, therefore the receiver may get a few bits garbage data after the frame end, we can use the remaining data length field to eliminate the garbage data as follows.

Suppose a data frame is split into three sub-frames, in the first sub-frame, the frame length field records the total frame length l_0 , the other two sub-frames record the remaining data length l_1 and l_2 , as depicted in Figure 6. When the first sub-frame comes, the receiver records the total length l_0 and writes the data into a buffer, when the second sub-frame comes, it extracts the remaining data length l_1 from the header and adjusts the buffer write pointer to l_0-l_1 , then continue to write the second sub-frame data into the buffer, so l_0-l_1 is the first sub-frame data length, l_0-l_2 is the first and second sub-frame length. When the received data length is equal to l_0 during receiving the third sub-frame, the receiver knows all frame data has been merged and finishes current frame. If there are errors or frame collisions, the received frame cannot pass CRC check and will be dropped.

There are an extreme situations during frame split and merge, as depicted in Figure 7. Suppose AP and T1 are transmitting data to each other, T1 firstly finishes and waits for AP to return an ACK, at the same time AP is very close to finish transmitting its frame, such as only one bit remains to the frame end, according to FD-CSMA/CA, AP should immediately stop sending and split the

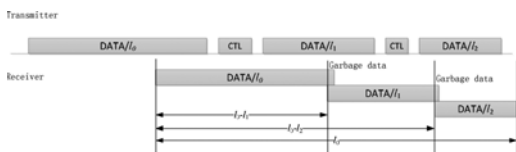


Figure 6. Eliminate the garbage data using remaining data length.

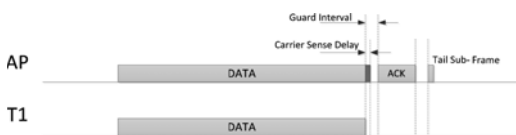


Figure 7. The Minimum tail sub-frame problem.

data, so there will be a very short tail sub-frame to send after ACK, we denote its transmitting time as t_{min} . The receiver T1 will continue receiving data until the channel is clear, assume the carrier sense delay is t_{cs} , if $t_{cs} \geq t_{min}$, the receiver will mistakenly think it has reached the frame end and begin to wait for AP to return an ACK. In fact, the received garbage data during carrier sense delay overwrites the short tail sub-frame, we call this the minimum tail sub-frame problem.

To solve the problem, obviously t_{cs} should less than t_{min} , so when transmitter need split frames, it firstly check the remaining data length, if its transmitting time is less than or equal to the t_{cs} , it will continue transmitting until finished. To ensure the FD-CSMA/CA work correctly, the CTS/ACK timeout interval should increase t_{cs} , which is very easier to implement.

3.3 Control frame split

We have analyzed how to split and merge the data frames, how about control frames? In fact splitting control frames should be avoided. On one hand, control frames are usually very short, splitting them will bring high overhead because each split sub-frame should append a physical layer header before transmitting, the physical layer header maybe longer than the control frame; on the other hand, some control frames such CTS/ACK must be returned in a very short interval, otherwise the receiver will timeout, splitting them will complicate the timeout mechanism.

But do not allow splitting control frames may cause control frame collisions. In FD-CSMA/CS protocol there are four types of control frames: Down link ACK (DACK), Downlink CTS (DCTS), Uplink RTS (URTS), Uplink ACK (UACK). Collisions only occur between the same direction control frames. Because a terminal can only wait for DCTS or DACK, never both, the AP has no chance to send them at the same time, so no collisions can occur between DCTS and DACK. Then considering URTS and UACK, if AP finishes transmitting a data frame during T1 is transmitting RTS, a collision may occur between URTS and UACK, as depicted in Figure 8.

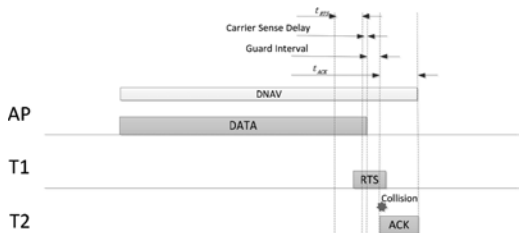


Figure 8. Collisions between URTS and UACK.

URTS and UACK collide because there is no synchronization between uplink and downlink channel, we use a virtual carrier sense mechanism to solve the problem. Each node maintains a Uplink Network Allocation Vector (UNAV) and Downlink Network Allocation Vector (DNAV) to indicate whether the uplink or downlink channel is busy. UNAV is updated according to the RTS/CTS duration field, if $UNAV > 0$, the terminals will freeze their backoff counters and not contend the channel. DNAV is updated according to the DATA duration field in the physical header. To avoid collisions between URTS and UACK, terminals should not contend channel when $DNAV > t_{RTS} + t_{cs} + t_{GI} + t_{ACK}$, so they must freeze their backoff counters if $UNAV > 0$ or $DNAV > t_{RTS} + t_{cs} + t_{GI} + t_{ACK}$.

4 TESTBED DESIGN AND IMPLEMENTATION

4.1 Hardware

We build a testbed for hybrid visible light communication networks using commodity components. Each node in the testbed consists of a BeagleBone Black board and a full duplex visible light transmitter and receiver.

BeagleBone Black board is a low cost microcomputer development platform, equipped with the TI AM3359 CPU (1GHz), 512MB DDR3 RAM and 4GB on-board flash storage, and provides 4 timers and 65 General-Purpose Input/Output (GPIO) pins. We use different color LEDs as uplink and downlink light source to support full duplex communication, such as red light for the downlink and green light for uplink. We use the same color LED as detectors instead of photodiode, because LED is an optical pass-band filter, its receiving bandwidth is very narrow, almost matches the one as LED transmitter, while the PD is a wide-band receiver that collects most of the light from the optical frequencies emitted by the sun, using LED as detector is a cost effective solution to implement low interference full duplex visible light channels.

We extend the BeagleBone Black board GPIO interface to connect to the optical transceiver: the GPIO output signal is amplified by a NPN transistor to drive the LED transmitter, the LED detector signal is firstly amplified by a transimpedance amplifier, then feed to an analog to digital converter, the conversion results are finally read by CPU from GPIO interface, as shown in Figure 9.

4.2 Software

The Linux operating system running within the BBB board is Debian 7.4 with Xenomai real time kernel. We implement FD-CSMA/CA MAC pro-

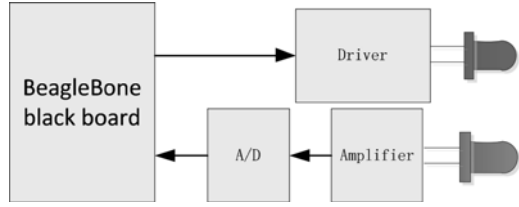


Figure 9. Full duplex VLC testbed.



Figure 10. Frame format of FD-CSMA/CA protocol.

col and some physical layer functions as a linux kernel module.

We use OOK modulation scheme in the physic layer, “HIGH” symbol means LED light on and “LOW” symbol means LED light off. Before modulation, frames are encoded with the manchester code, bit 1 is mapped to symbol sequence LOW-HIGH, and bit 0 is mapped to HIGH-LOW. Demodulation is performed with direct detection based on the received signal’s voltage.

There are four types of Frames in FD-CSMA/CA protocol: DATA/RTS/CTS/ACK, the frame format are showed in Figure 10. When a MAC layer DATA frame (MAC HDR and MAC DATA) is split into multiple sub-frames, each sub-frames has its own PHY HDR, they are transmitted as independent physical frames. The Split field indicates whether this is the first sub-frame of a MAC frame.

The carrier sensing and backoff process is similar to CSMA/CA protocol, the FD-CSMA/CA implementation is connected with the TCP/IP layers of the Linux operating system and thus the MAC protocol becomes transparent to various applications.

5 PERFORMANCE EVALUATION

We use iperf to test network protocol throughput, iperf is a widely used tool for active measurements of the maximum achievable bandwidth on IP networks. Iperf supports two test modes: single direction mode and dual direction mode. In single direction mode it only generates single direction traffic, such as from client to server. In dual direction mode, the client and server generate traffic to each other simultaneously, it is suitable to test full duplex link performance. We use both modes in our tests.

To compare the protocol performance we also implement the CSMA/CA protocol on our testbed.

In CSMA/CA the AP and terminals treat the uplink and downlink as one logical channel, transmitting on the uplink and downlink at the same time will cause a collision. The symbols LOW and HIGH interval are both set to 400 μ s, so the maximum physical data rate is 1250 bps. We assume one slot time equals 40 symbol time. The initial uplink backoff window sets to 16, backoff is not needed for downlink in FD-CSMA/CA because there is no channel contention for AP.

We first evaluate the MAC layer throughput of a two-node network. We use iperf dual testing mode to generate two UDP flows on uplink and downlink channel simultaneously, both nodes are saturated with fix payload packets. The saturation throughput is shown in Figure 11, where the payload varies from 50 to 400 bytes. With the payload size increase, both the throughputs of CSMA/CA and FD-CSMA/CA also increase. At all payload size, the FD-CSMA/CA throughput is improved dramatically compared to the CSMA/CA protocol, e.g. if payload is 50 bytes, the CSMA/CA throughput is 244 bps, while FD-CSMA/CA throughput

is 710 bps; if payload size is 400 bytes, the CSMA/CA throughput is 927 bps, while FD-CSMA/CA throughput is 1895 bps. Figure 13 also shows the FD-CSMA/CA uplink and downlink throughputs separately, the uplink throughput is lower than downlink because of two reasons, first is the RTS/CTS handshake consume some uplink bandwidth, second is the AP need not back-off while the uplink back-off period waste some bandwidth.

We also test the throughputs of CSMA/CA under single direction traffic and dual direction traffic. In CSMA/CA protocol at any time only one direction can transmit frames, so both the AP and terminal must back-off to contend the channel, under dual direction traffic there are collisions while in FD-CSMA/CA there aren't. We can see under single direction test the CSMA/CA outperforms the uplink channel throughput of FD-CSMA/CA because FD-CSMA/CA has additional frame split overheads. While under dual direction test, the CSMA/CA throughput is very close to the uplink channel of FD-CSMA/CA, the reason is the CSMA/CA collision overheads compensate the frame split overheads of FD-CSMA/CA.

Then we test the split and merge overhead of FD-CSMA protocol. We generate a single UDP flow for uplink and downlink separately, in this situation there are no frame splits at all. Obviously the throughput of each direction in single direction test should be higher than that in dual direction test, Figure 12 shows the results. The split overhead is about 10% for uplink and 13% for downlink, this is the cost to transmit simultaneously on uplink and downlink channels. Figure 12 also shows the total throughput of two single direction traffics, although it is higher than dual total throughput, it is only a theoretical throughput and cannot be obtained in practical environment.

Next we test the performance of a three-node network, as shown in Figure 13. One node acts as a base station and the other two as clients,

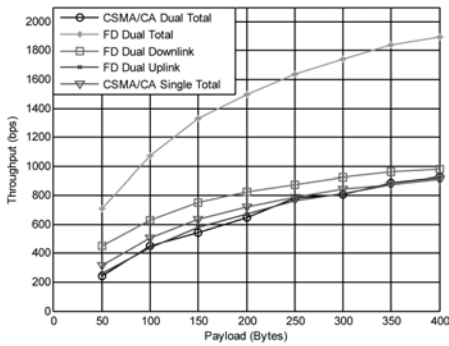


Figure 11. Performance comparison of CSMA/CA and FD-CSMA/CA.

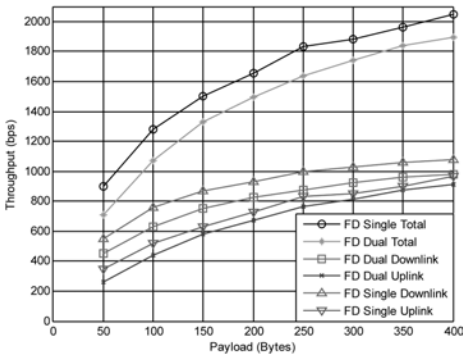


Figure 12. FD-CSMA/CA split overhead.

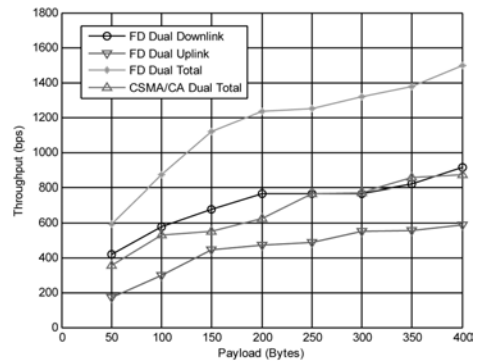


Figure 13. Performance of a three-node network.

this scenario emulates a light bulb in the ceiling and two users under the light. The throughput increase with various payloads is similar to the two-node network. The FD-CSMA/CA outperforms the CSMA/CA at all payloads as expected, the average throughput improvement is 75% compared to CSMA/CA. The Uplink throughput of FD-CSMA/CA is much lower than the downlink throughput, besides the reasons we analyze in the two-node network, another reason is there are no channel contentions on the downlink channel while there are two nodes contending the uplink channel, collisions often occur under saturation conditions which further degrade the uplink throughput.

6 CONCLUSION

In this paper, we presented the design, implementation, and performance evaluation of the FD-CSMA/CA protocol for full duplex hybrid visible light communication networks. By dynamically splitting and merging MAC frames, data can be transmitted concurrently on uplink and downlink channel. We built a VLC testbed and implement the FD-CSMA/CA protocol in Linux network stack. Performance evaluation shows its effectiveness to improve network throughput.

ACKNOWLEDGEMENT

The work reported in this paper was supported in part by the National Basic Research Program of China (973 Program) under grant 2012 CB933504 in China.

REFERENCES

- [1] IEEE standard for local and metropolitan area networks—part 15.7: Short-range wireless optical communication using visible light. IEEE Std 802.15.7–2011 (Sept 2011), 1–309.
- [2] P.H. Pathak, X. Feng, P. Hu, P. Mohapatra. Visible Light Communication, Networking and Sensing: A Survey, Potential and Challenges. IEEE Communications Surveys and Tutorials, 2015.
- [3] Bharadia, D., McMilin, E., and Katti, S. Full duplex radios. In Proceedings of the ACM SIGCOMM (2013), pp. 375–386.
- [4] Choi, J.I., Jain, M., Srinivasan, K., Levis, P., and Katti, S. Achieving single channel, full duplex wireless communication. In Proceedings of the ACM MobiCom (2010), pp. 1–12.
- [5] Dietz, P., Yerazunis, W., and Leigh, D. Very low-cost sensing and communication using bidirectional LEDs. In TR2003–35 (2003).

- [6] Giustiniano, D., Tippenhauer, N., and Mangold, S. Low-complexity visible light networking with LED-to-LED communication. In Proceedings of the IFIP Wireless Days (WD) (2012), pp. 1–8.
- [7] OpenVLC. <http://openvlc.org/openvlc.html>.
- [8] BeagleBone Black. <http://beagleboard.org/Products/BeagleBone+Black>.
- [9] Jain, M., Choi, J.I., Kim, T., Bharadia, D., Seth, S., Srinivasan, K., Levis, P., Katti, S., and Sinha, P. Practical, real-time, full duplex wireless. In Proceedings of the ACM MobiCom (2011), pp. 301–312.
- [10] Lin, K., and Hirohashi, K. High-speed full-duplex multiaccess system for LEDs based wireless communications using visible light. In Proceedings of the International Symposium on Optical Engineering and Photonic Technology (OEPT) (2009), pp. 1–5.
- [11] Liu, C. B., Sadeghi, B., and Knightly, E. W. Enabling vehicular visible light communication (V2LC) networks. In Proceedings of the ACM VANET (2011), pp. 41–50.
- [12] Schmid, S., Corbellini, G., Mangold, S., and Gross, T. R. LED-to-LED visible light communication networks. In Proceedings of the ACM MobiHoc (2013), pp. 1–10.
- [13] Wang, Q., Giustiniano, D., and Puccinelli, D. OpenVLC: Software-defined visible light embedded networks. In 1st ACM Workshop on Visible Light Communication Systems, in conjunction with MobiCom 2014 (September 2014), pp. 1–6.
- [14] Q. Wang, D. Giustiniano, “Communication Networks of Visible Light Emitting Diodes with Intra-Frame Bidirectional Transmission”, Proceedings of the 10th ACM International on Conference on emerging Networking Experiments and Technologies, Pages 21–28, ACM New York, NY, USA, 2014.
- [15] L. Zhang, X.H. Wang, A Distributed Broadcast Algorithm Based on Master/Slave Dominators for Wireless Mesh Networks, The 2nd International Conference on Mechanical Engineering, Industrial Electronics and Information Technology, pp. 2217–2222, Chongqing, Sep. 2013.
- [16] Li Li, Bin Qin, Chunyuan Zhang. Efficient broadcasting in multi-radio multi-channel and multi-hop wireless networks based on self-pruning. HPCC’07 Proceedings of the Third international conference on High Performance Computing and Communications. pp. 484–495, 2007.
- [17] Lei Zhang, Xuehui Wang. DCF Fairness Enhancement Algorithm for Multi-hop Wireless Ad Hoc Networks. The 7th International Conference on Wireless Communications, Networking and Mobile Computing, 2011.
- [18] D. Giustiniano, N. Tippenhauer, and S. Mangold. Low-Complexity Visible Light Networking with LED-to-LED Communication. IFIP Wireless Days 2012, Nov. 2012.
- [19] D. O’Brien. Visible Light Communications: Challenges and Potential. In Photonics Conference (PHO), 2011 IEEE, pages 365–366, Oct. 2011.
- [20] J. Zhang, X. Zhang, and G. Wu. Dancing with light: Predictive in-frame rate selection. In Proc. of the IEEE INFOCOM, pages 1–9, 2015.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Combining feature extraction methods to classify three motor imagery tasks

M.H. Zaky, A.A. Nasser & M.E. Khedr

Electronics and Communications Engineering Department, AASTMT, Alexandria, Egypt

ABSTRACT: In Brain Computer Interface (BCI), a subject's thoughts are read to provide an appropriate way of communication where only brain signals are used. The Information of Electroencephalogram (EEG) signals defer between subjects depending on their thoughts according to research. In this paper, a combined feature extraction method is proposed to obtain the best result of discrimination between three Motor Imagery (MI) movements which are Left Hand, Right Hand and both Feet through an offline analysis for two subjects. The signals are featured using statistical time domain features, Band Power (BP), statistical from Discrete Wavelet Transform (DWT) coefficients and Common Spatial Pattern (CSP), then the features are classified using Linear Discriminant Analysis (LDA), K-Nearest Neighbors (K-NN), Naive Bayes and Linear Support Vector Machine (SVM). The combination of CSP and statistical time domain features classified using LDA is found to outperform all other combinations with an average accuracy above 95%.

1 INTRODUCTION

Recently EEG based BCI duo to its low cost and convenient measurement has become a very interesting topic in the signal processing research field (Major & Conrad, 2014). BCI's main aim is to develop a new communication channel between the human brain and the computer without using any peripheral nervous system or muscles. This function involves the classification of different brain states with a high accuracy requirement which is the main goal in BCI research. This system which is a Non-invasive system is composed of many parts (Vargic, et al., 2015) as shown in Figure 1.

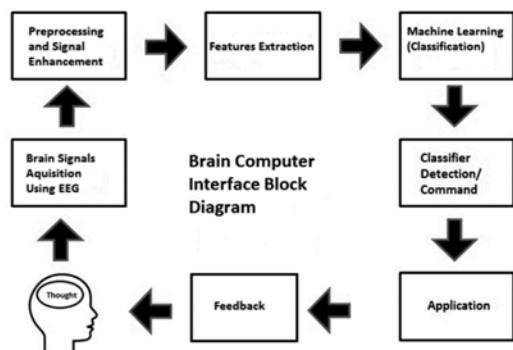


Figure 1. BCI system block diagram.

Human body or imagery movement can cause a change in the neuronal activity in the primary sensorimotor areas. A lot of oscillations originate from the sensory motor cortex. Event-Related Synchronization (ERS) in the gamma (γ) band and Event-Related Desynchronization (ERD) in the mu (μ) and beta (β) bands of the EEG develops in the brain during imagination or body-part movement. The γ ERS and the μ - β ERD occur at the contralateral side of the brain near the somatosensory and motor cortex areas during particular limb movement. In case of ERS the power of the γ component increases, while in case of ERD the power of the μ - β component decreases. However BCI research is relevant only to the μ rhythm (8–13 Hz) and the central β rhythm (13–30 Hz) (Stankevich & Spitsyn, 2015, Duan, et al., 2014).

Reference (Mina, et al., 2006) tried to extract statistical features from EEG signal in time domain and used K-NN for classification, while in (Gaur, et al., 2015) BP feature was extracted and classified using LDA classifier and in (Azalan, et al., 2014) Neural Networks classifier was used, many statistical features were extracted from DWT coefficients like max, variance, energy and entropy and classified using K-NN in (Imran, et al., 2014) and also classified using LDA and SVM in (Carrera-Leon, et al., 2012). In addition DWT was taken as feature and fed to LDA, K-NN, Naïve Bayes and SVM classifiers in (Bhattacharyya, et al., 2011), also the CSP as a feature has proven its efficiency as mentioned in (Xie & Li, 2015) where it was classified

using SVM and in (Wu, et al., 2013) where LDA was used for classification.

In this paper, a report on the offline analysis of a three class Motor Imagery (MI) EEG based BCI experiments is done to propose a combined feature extraction method for synchronous three class MI based BCI experiment. In MI the subject is required to imagine a specific movement then the EEG signals corresponding to this movement will be classified in order to translate it to a control signal. A variety of combined features are tested with four different classifiers. Time statistical features, statistical features from DWT coefficients, CSP and BP distributions of ERD/ERS are chosen as features. The features are taken as inputs to classifiers (LDA, Linear SVM, Naïve Bayes and K-NN) to be classified into one of three classes Left Hand (LH), Right Hand (RH) or both Feet (F). Classification Accuracy and feature vector length are used to evaluate performance of each approach.

2 METHODOLOGY AND IMPLEMENTATION

2.1 Dataset

Datasets are provided by the Dr. Cichocki's Lab (Lab. for Advanced Brain Signal Processing), BSI, RIKEN collaboration with Shanghai Jiao Tong University (Qibin Zhao Research Scientist 2015). These datasets of EEG were recorded from two healthy subjects. The cue-based BCI paradigm consisted of three MI tasks, namely the imagination of movement of LH, RH and F. Each subject was sitting in a comfortable armchair in front of a computer screen. At the beginning of a trial, the screen is blank. After two seconds ($t = 2s$), a cue in the form of an arrow pointing either to the left, right or down (corresponding to three classes of LH, RH and F) appeared and stayed on the screen for a specific duration (3–10 sec). This prompted each subject to perform the desired MI task. Each subject was requested to carry out the MI task until the cue disappeared from the screen and try to avoid the eye blinking or eye movements during the imagination. A 2 seconds break followed when the cue is disappeared. This procedure was repeated 30–100 times for each run with the random cue sequence. The paradigm is illustrated in Figure 2.

2.2 Preprocessing

In this data set, the device g.tec (g.USBamp) was used for recording the EEG signals. The EEG signals were band-pass filtered between 2Hz and 30Hz with sample rate of 256Hz and a notch filter at 50Hz was enabled for g.tec device, the signals

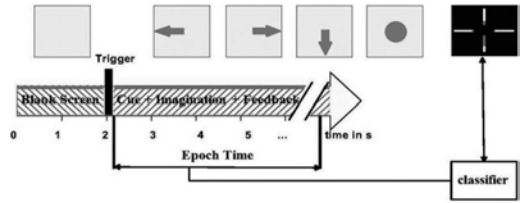


Figure 2. Timing scheme for the paradigm.

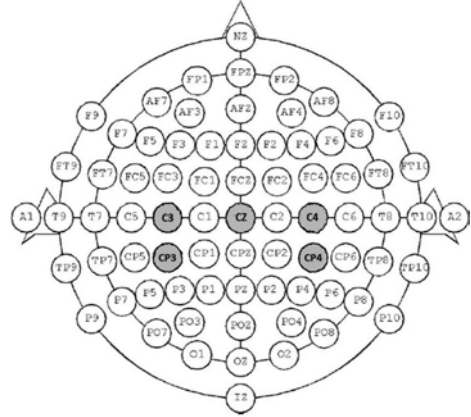


Figure 3. Electrodes positions.

were measured in V. The electrodes montage is shown in Figure 3, the electrodes of C3, CP3, C4, CP4, CZ were used to record the EEG signals.

Prior to the feature extraction step, the recorded EEG signals are bandpass filtered between 8–30 Hz using a 2nd order elliptic filter to extract the band of frequencies related to the MI Signals.

3 FEATURE EXTRACTION

For all datasets, a feature is extracted from each channel alone and normalized to form a feature vector then all feature vectors of all channels are combined together to form the feature vector of one sample, then all the samples are combined to form the feature matrix of the dataset which size is $N \times M$, where N = the number of samples per session and M = the total feature vector length of all channels combined, the following is the description of each feature used.

3.1 Statistical time domain

The standard deviation is calculated using Equation (1), also the variance using Equation (2) on the EEG signals in time domain (Rak, et al., 2012).

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (1)$$

$$\sigma^2 = \frac{1}{2} \sum_{i=1}^N (X_i - \mu)^2 \quad (2)$$

where σ = standard deviation, σ^2 = variance, μ = mean of all values of the data recorded from one channel and N = the number of values.

3.2 Band Power

For the filtered band between 8–30 Hz, the BP is calculated for each channel using Equation (3) where N = the number of values (Aydemir, et al., 2011).

$$BP = \frac{|abs(FFT(Signal))|^2}{N} \quad (3)$$

3.3 Discrete Wavelet Transform

DWT is a multiresolution technique that decomposes the signal into Approximation coefficients (A) and Detail coefficients (D). The approximation coefficients are then divided into new approximation coefficients and detail coefficients. This process is performed iteratively, producing a set of approximation coefficients and detail coefficients according to the number of decomposition levels. It exhibits a good frequency resolution at low frequencies and a good time resolution at high frequencies so it gives a precise time-frequency information about the signal, also it has a low computational cost and easy to implement (Imran, et al., 2014) and (Carrera-Leon, et al., 2012).

Using db4 as mother wavelet and by applying only one level of decomposition, four statistical features (Max, Standard Deviation, Variance and Entropy) are taken from the coefficients to reduce the high dimensionality.

The DWT coefficients of a signal $x[n]$ can be obtained using Equation (4).

$$C(a,b) = \sum_{n \in \mathbb{Z}} x[n] \Psi_{a,b}[n] \quad (4)$$

where a = dilation or scale, b = translation, and $\Psi_{a,b}[n]$ is the discrete wavelet which is expressed as Equation (5).

$$\Psi_{a,b}(n) = \left(\frac{1}{\sqrt{a}} \right) \times \Psi \left(\frac{n-b}{a} \right) \quad (5)$$

3.4 Common Spatial Pattern

CSP algorithm can be used to extract the task's related signal component and eliminate the task's

unrelated signal component and noises of various types using a linear transform to project the multi-channel EEG data into a low-dimensional spatial subspace thus it maximizes the variance of the two-class signal matrices (Xie & Li, 2015, Wu, et al., 2013).

Basically it's applied to binary classification problems, here One V.s All technique is used to apply CSP to three class classification. The basic algorithm is illustrated as follows for a binary case.

X_H and X_F = the preprocessed two classes' samples (hand and foot) respectively with dimensions $N \times T$, where N = the number of electrodes and T = the number of sampling points of each electrode. The normalized spatial covariance of the EEG data can be represented as R_H and R_F as in Equation (6).

$$R_F = \frac{X_F X_F^T}{\text{trace}(X_F X_F^T)} \quad R_H = \frac{X_H X_H^T}{\text{trace}(X_H X_H^T)} \quad (6)$$

The averaged normalized covariance \bar{R}_H and \bar{R}_F are calculated by averaging over all the trails of each group. The composite spatial covariance is the sum of \bar{R}_H and \bar{R}_F as in Equation (7).

$$R = \bar{R}_F + \bar{R}_H = U_0 \Sigma U_0^T \quad (7)$$

where U_0 = the matrix of eigenvectors and Σ = the diagonal matrix of eigenvalues of U_0 . The whitening transformation matrix is obtained by Equation (8).

$$P = \Sigma^{-1/2} U_0^T \quad (8)$$

Then the average covariance matrices can be transformed to Equations (9).

$$S_F = P \bar{R}_F P^T \quad S_H = P \bar{R}_H P^T \quad (9)$$

S_H and S_F share common eigenvectors and the sum of corresponding eigenvalues for the two matrices will always be one, refer to Equations (10).

$$S_H = U \Sigma_H U^T \quad S_F = U \Sigma_F U^T \quad \Sigma_H + \Sigma_F = I \quad (10)$$

The eigenvectors with the largest eigenvalues for S_H have the smallest eigenvalues for S_F and vice versa. The transformation of whitened EEG onto the eigenvectors corresponding to the largest eigenvalues in Σ_H and Σ_F is optimal for separating variance into two signal matrices. The projection matrix W is denoted as Equation (11).

$$W = U^T P \quad (11)$$

With the projection matrix W , the original EEG can be transformed into uncorrelated components using Equation (12).

$$Z = WX \quad (12)$$

Z can be seen as EEG source components including common and specific components of different tasks. Based on the projected signal trials, the classifier is trained on the feature vectors obtained by normalizing and log-transforming the variances of projected EEG series as Equation (13).

$$f_k = \log \left(\frac{\text{diag}(Z_k Z_k^T)}{\text{trace}(Z_k Z_k^T)} \right) \quad (13)$$

where f_k = the projected CSP feature vector of the k -th EEG sample. The log transformation serves to approximate normal distribution of the data.

4 CLASSIFICATION

Each feature matrix is divided into (75%) of samples for training and (25%) for testing, the results of each combination (feature/s and classifier) are recorded 100 time and each run is splitted according to the ratio mentioned in a random order, then the results are sorted in a descending order and the top ten are averaged to get the final estimated result which is recorded in each classifier table.

4.1 Linear Discriminant Analysis

LDA is a classification method originally developed in 1936 by R. A. Fisher, it reduce the feature space while keeping much of the class discriminatory information. This is done by projecting the data into a low-dimensional feature space with maximization of inner and intra class distances, this achieves maximum discrimination between-classes. LDA is simple, mathematically robust and often produces models with a good accuracy (Var- gic, et al., 2015).

Suppose, we need to discriminate between the two classes B1 and B2. Then, we classify the d -dimensional sample points $x = \{x_1, x_2, \dots, x_d\}$, so our linear combination Equation will be (14).

$$Z = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_d X_d \quad (14)$$

The objective of LDA is to find a linear combination of variables (predictors) that gives

maximum class separability. To reach this, Fisher defined the following score function denoted as Equation (15).

$$S(\beta) = \frac{\beta^T \mu_1 - \beta^T \mu_2}{\beta^T C \beta} \quad (15)$$

where β = the linear model coefficients vector which is $\{\beta_1, \beta_2, \dots, \beta_d\}$, μ_1 and μ_2 = the mean vectors for class 1 subset and class 2 subset respectively and C = the pooled covariance matrix.

In this method, we need to estimate the linear coefficients that maximize the score function to maximize the discrimination between the two considered classes. We can find β and C using these Equations (16) and (17).

$$\beta = C^{-1}(\mu_1 - \mu_2) \quad (16)$$

$$C = \frac{1}{n_1 + n_2}(n_1 C_1 + n_2 C_2) \quad (17)$$

where C_1 and C_2 = the covariance matrices for class 1 subset and class 2 subset respectively and n_1 and n_2 = the number of observations in class 1 and class 2 respectively.

We need to measure the separation between the two classes chosen for assessing the effectiveness of the discrimination, and one way to do that is to calculate the Mahalanobis distance between two groups (Gaur, et al., 2015) using Equation (18).

$$\Delta^2 = \beta^T (\mu_1 - \mu_2) \quad (18)$$

where Δ = the Mahalanobis distance between two groups.

4.2 Support Vector Machine

SVM performs classification by separating features of different classes linearly using a clear gap that is as wide as possible, and those vectors that define the hyperplane are called support vectors. SVM can deal with large feature vectors as the dimensionality doesn't affect the complexity of the algorithm. It's sensitive to noise and can consider only two classes so the One V.s All technique is used to solve the three class problem. The limitations of speed and size (in both training and testing) depend on the kernel function chosen, so linear kernel function is used to make it as simple as possible (Bhattacharyya, et al., 2011).

In linear SVM the linear decision boundaries are used. For example in this linear separable training sample $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, it finds the optimal classification plane that separates the two

class sample, where the interval is maximum, in order to satisfy the constraint conditions in Equation (19) (Hong, et al., 2015):

$$\begin{cases} (x_i \cdot w) + b \geq 1, y_i = 1 \\ (x_i \cdot w) + b \leq -1, y_i = -1 \end{cases} \quad (19)$$

where x_i represents the feature vector, y_i represents the category number, w denotes the projection vector of classification hyperplane and b denotes classification threshold. In this condition, the classification function is represented by Equation (20).

$$f(x) = \text{sgn}(w \cdot x + b) \quad (20)$$

4.3 Naïve Bayes

The Naive Bayesian classifier is based on Bayes' theorem. This classifier is useful for large datasets as its model is simple to build because of the simplicity of its iterative parameter estimation. Bayes theorem idea is about calculating the posterior probability $P(c|x)$ using Equation (21), from class prior probability $P(c)$, predictor prior probability $P(x)$, and likelihood $P(x|c)$ which is the probability of predictor given class (Bhattacharyya, et al., 2011),

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (21)$$

This classifier assume that the effect of the value of any predictor on a given class is independent of the values of other predictors and hence it solve the problem of the complexity of calculating the joint probability distribution, and this is called Naive Bayesian assumption (Bhattacharyya, et al., 2011) and hence $P(c|x)$ can be calculated using Equation (22). Normal Gaussian distribution is used in building the classifier model used in this paper.

$$P(c|x) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c) \quad (22)$$

4.4 K-Nearest Neighbours

K-NN is one of the simplest machine learning algorithms, it stores all available cases and classifies the new ones based on a similarity measure like distance functions. Depending on the majority of votes of its neighbors a new case is classified, it's assigned to the class that is most common among its K nearest neighbors measured by the distance

function and Euclidean distance is used to build the model. If $K = 1$ as what it is used here, then the new case will be assigned to the class of its nearest neighbor (Aydemir & Kayikcioglu, 2011). Equation (23) shows the calculation of Euclidean Distance.

$$D_E(a, b) = \sum_{i=1}^N \sqrt{(a_i - b_i)^2} \quad (23)$$

where a and b = the training and testing EEG signals composed of N features respectively.

K-NN algorithm has many defects that makes it not popular in this field, for example it's sensitive to noise and is computationally expensive. K-NN is nearly optimal for large sample limits also they may be efficient with low dimensional feature vectors (Vargic, et al., 2015).

5 RESULTS AND DISCUSSION

The sessions have different number of samples, and the number of samples of each session is divided equally over the three tasks, the number of samples for each subject's session can be found in Table 1.

As illustrated in the following tables, four types of classifiers are used to test how beneficial each feature vector is either separate or combined for EEG signals of subject A and Subject C, and four types of features are tested.

Classification accuracy is calculated to evaluate the performance according to this Equation (24).

$$Accuracy = \left(\frac{N_{correct}}{N_{total}} \right) \times 100\% \quad (24)$$

where N_{total} = number of overall samples to be classified, and $N_{correct}$ = number of correct classified samples.

The performance of each combination (feature/s and classifier) is evaluated using two parameters: the average classification accuracy and the feature vector length. According to these parameters statistical features from time domain combined with CSP feature gives the highest classification accuracy with all classifiers used, also its vector length is only 25 so the computational time is low, the vector length of each feature alone is shown in Table 6,

Table 1. Number of samples for each subject.

Subject	Subject A	Subject C
Samples	270	180

Table 2. LDA classifier results.

Feature	Subject A	Subject C	Average
F2	92.4	90.96	91.68
F1	93.15	92.49	92.82
F4	94.73	92.67	93.7
F1+F2	92.38	94.62	93.5
F3	92.73	92.99	92.86
F2+F4	96.84	92.61	94.73
F1+F4	96.2	94.24	95.22
F2+F3	93.32	90.97	92.15
F1+F3	88.74	91.07	89.91
F1+F2+F4	96.15	91.35	93.75
F3+F4	96.08	92.37	94.23
F1+F2+F3	90.09	88.93	89.51
F2+F3+F4	96.06	90.96	93.51
F1+F3+F4	95.56	90.21	92.89
All	94.24	87.81	91.03

Table 3. SVM classifier results.

Feature	Subject A	Subject C	Average
F2	84.45	90.31	87.38
F1	88.08	91.21	89.65
F4	91.15	91.96	91.56
F1+F2	87.39	90.97	89.18
F3	88.07	91.77	89.92
F2+F4	93.83	91.97	92.9
F1+F4	95.11	93.46	94.29
F2+F3	86.14	90.63	88.39
F1+F3	86.81	91.39	89.1
F1+F2+F4	94.5	91.58	93.04
F3+F4	93.94	91.31	92.63
F1+F2+F3	88.65	91.52	90.09
F2+F3+F4	94.74	93.36	94.05
F1+F3+F4	94.29	91.82	93.06
All	94.11	91.13	92.62

and the feature vector length for any combination of features can be found by adding these lengths.

For LDA classifier when using statistical features from time domain combined with CSP as a feature, an average classification accuracy of 95.22% can be obtained which is the highest of all classifiers. By taking the average classification accuracy as the most important parameter for deciding the best combination, the SVM classifier comes in the second place by 94.29%. K-NN and Naïve Bayes classifiers are also used but they didn't give good results compared to SVM and LDA. Another thing is realized regarding the classification accuracy, LDA outperforms SVM with all features combination, but when the feature vector length is increased SVM will be better (Lee, et al.,

Table 4. KNN classifier results.

Feature	Subject A	Subject C	Average
F2	86.07	84.58	85.33
F1	86.22	83.88	85.05
F4	88.99	82.91	85.95
F1+F2	87.39	85.84	86.62
F3	83.17	84.47	83.82
F2+F4	89.34	83.2	86.27
F1+F4	87.44	88.32	87.88
F2+F3	84.34	83.36	83.85
F1+F3	83.01	82.92	82.97
F1+F2+F4	88.34	87.29	87.82
F3+F4	87.18	88.14	87.66
F1+F2+F3	85.43	84.48	84.96
F2+F3+F4	87.59	87.49	87.54
F1+F3+F4	88.69	84.65	86.67
All	88.51	86.46	87.49

Table 5. Naïve Bayes classifier results.

Feature	Subject A	Subject C	Average
F2	76.74	84.33	80.54
F1	81.12	82.29	81.71
F4	84.3	90.05	87.18
F1+F2	79.36	83.96	81.66
F3	78.04	82.74	80.39
F2+F4	87.8	91.3	89.55
F1+F4	88.94	90.34	89.64
F2+F3	76.26	83.02	79.64
F1+F3	79.63	85	82.32
F1+F2+F4	86.54	90.77	88.66
F3+F4	85.85	89.12	87.49
F1+F2+F3	78.63	81.72	80.18
F2+F3+F4	85.23	88.28	86.76
F1+F3+F4	84.8	87.87	86.34
All	86.42	86.52	86.47

2005), this can be realized in the last three feature vector combinations, also the results in most cases have proved that combining features is more effective than using one feature in improving the classification accuracy (Hong, et al., 2015).

6 CONCLUSION

This paper focused on feature extraction techniques because it's the precondition of classification, an offline three-class motor imagery based BCI has been implemented using different combinations of features and classifiers. Four types of classifiers are used to test different feature combinations composed of statistical features from time domain,

Table 6. Feature Vector Length.

Feature	Time (F1)	BP (F2)	DWT (F3)	CSP (F4)
Length	10	5	20	15

band power, statistical features from DWT coefficients and CSP. The experiments proved that statistical time domain features combined with CSP feature and classified by LDA achieves the best classification performance among all other combinations evaluated by two parameters, the average classification accuracy and feature vector length, so it was found to be the most robust model. In addition, the whole program included feature extraction and pattern recognition has been written by Matlab language. In the future we will work on optimizing both features and classifiers to decrease the computational time as an initial step to build a real time system.

REFERENCES

- Aydemir, O. and Kayikcioglu, T., 2011. Wavelet transform based classification of invasive brain computer interface data. *Radioengineering*, 20(1), pp. 31–38.
- Aydemir, O., Ozturk, M. and Kayikcioglu, T., 2011, August. Performance evaluation of five classification algorithms in low-dimensional feature vectors extracted from eeg signals. In *Telecommunications and Signal Processing (TSP), 2011 34th International Conference* on (pp. 403–407). IEEE.
- Azalan, M.S.Z., Paulraj, M.P. and Bin Yaacob, S., 2014, August. Classification of hand movement imagery tasks for brain machine interface using feed-forward network. In *Electronic Design (ICED), 2014 2nd International Conference* on (pp. 431–436). IEEE.
- Bhattacharyya, S., Khasnobish, A., Konar, A., Tibarewala, D.N. and Nagar, A.K., 2011, April. Performance analysis of left/right hand movement classification from EEG signal by intelligent algorithms. In *Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), 2011 IEEE Symposium* on (pp. 1–8). IEEE.
- Carrera-Leon, O., Ramirez, J.M., Alarcon-Aquino, V., Baker, M., D’Croze-Baron, D. and Gomez-Gil, P., 2012, May. A motor imagery BCI experiment using wavelet analysis and spatial patterns feature extraction. In *Engineering Applications (WEA), 2012 Workshop* on (pp. 1–6). IEEE.
- Duan, S., Xu, T., Zhuang, W. and Mao, D., 2014, June. The feature extraction of ERD/ERS signals based on the wavelet package and ICA. In *Intelligent Control and Automation (WCICA), 2014 11th World Congress* on (pp. 5621–5625). IEEE.
- Gaur, P., Pachori, R.B., Wang, H. and Prasad, G., 2015, July. An empirical mode decomposition based filtering method for classification of motor-imagery EEG signals for enhancing brain-computer interface. In *Neural Networks (IJCNN), 2015 International Joint Conference* on (pp. 1–7). IEEE.
- Hong, J., Qin, X., Bai, J., Zhang, P. and Cheng, Y., 2015, August. A combined feature extraction method for left-right hand motor imagery in BCI. In *Mechatronics and Automation (ICMA), 2015 IEEE International Conference* on (pp. 2621–2625). IEEE.
- Imran, S.M., Talukdar, M.T.F., Sakib, S.K., Pathan, N.S. and Fattah, S.A., 2014, April. Motor imagery EEG signal classification scheme based on wavelet domain statistical features. In *Electrical Engineering and Information & Communication Technology (ICEEICT), 2014 International Conference* on (pp. 1–4). IEEE.
- Lee, F., Scherer, R., Leeb, R., Neuper, C., Bischof, H. and Pfurtscheller, G., 2005. A comparative analysis of multi-class EEG classification for brain computer interface. In *Proceedings of the 10th Computer Vision Winter Workshop* (pp. 195–204).
- Major, T.C. and Conrad, J.M., 2014, March. A survey of brain computer interfaces and their applications. In *SOUTHEASTCON 2014, IEEE* (pp. 1–8). IEEE.
- Mina, R.T., Atiya, A., Owis, M.I. and Kadah, Y.M., 2006. Brain-Computer Interface Based on Classification of Statistical and Power Spectral Density Features. *Biomedical Engineering*, pp. 2–5.
- Qibin Zhao Research Scientist, Lab. for Advanced Brain Signal Processing, Brain Science Institute, RIKEN. 2–1, Hirosawa, Wakoshi, Saitama, Japan. Available From: <http://www.bsp.brain.riken.jp/~qibin/homepage/Home.html>. [15 September 2015].
- Rak, R.J., Kołodziej, M. and Majkowski, A., 2012. Brain-computer interface as measurement and control system the review paper. *Metrology and Measurement Systems*, 19(3), pp. 427–444.
- Stankevich, P. and Spitsyn, V.G., 2015, July. A review of Brain-Computer Interface technology. In *Institute of Electrical and Electronics Engineers Inc*.
- Vargic, R., Chlebo, M. and Kacur, J., 2015, September. Human computer interaction using BCI based on sensorimotor rhythm. In *Intelligent Engineering Systems (INES), 2015 IEEE 19th International Conference* on (pp. 91–95). IEEE.
- Vijean, V., Hariharan, M., Saidatul, A. and Yaacob, S., 2011, October. Mental tasks classifications using S-transform for BCI applications. In *Sustainable Utilization and Development in Engineering and Technology (STUDENT), 2011 IEEE Conference* on (pp. 69–73). IEEE.
- Wu, S.L., Wu, C.W., Pal, N.R., Chen, C.Y., Chen, S.A. and Lin, C.T., 2013, April. Common spatial pattern and linear discriminant analysis for motor imagery classification. In *Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), 2013 IEEE Symposium* on (pp. 146–151). IEEE.
- Xie, Y. and Li, X., 2015, October. A brain controlled wheelchair based on common spatial pattern. In *Bioelectronics and Bioinformatics (ISBB), 2015 International Symposium* on (pp. 19–22). IEEE.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Telemedicine home program in patients with cystic fibrosis: Results after 10 Years

F. Murgia, I. Tagliente, V. Mercuri & S. Bella

Department of Special Pediatrics, Bambino Gesù Pediatric Hospital–IRCCS, Rome, Italy

F. Bella

Department of Computer Science, University “La Sapienza”, Rome, Italy

I. Zoppis, G. Mauri & F. Sicurello

Department of Computer Science, University of “Milano-Bicocca”, Milano, Italy

ABSTRACT: Objectives: We studied the effect of Telehomecare (THC) in a group of Cystic Fibrosis (CF) patients. Specifically, in order to investigate the evolution of the clinical trend, we examined the monitoring activities of CF patients followed at home for a period of 4.5 years.

Methods: Forced Expiratory Volume in the first second (FEV1) was monitored at home, with the aim of an early recognition of the relapses of pulmonary infections. FEV1 was monitored by Spirotek-MIR instrumentation for 4.5 years, using THC as a tool, in addition to the standard therapeutic protocol. 16 CF patients were followed by doctors specialized in the treatment, over a period of 4.5 years. As control group, we enrolled 16 CF patients with similar characteristics of age, degree of pulmonary involvement, bacterial colonization and O2 dependency.

Results: While in the first 5 year report we showed statistically significant reduction in hospital admissions and a tendency over time towards a better stability of the respiratory function, here we provide an evidence of the significant different evolution between THCs and controls measured as the difference between the values of average annual FEV1 and those values observed during a starting reference year (i.e. 2011).

Discussion: The obtained results in both quantitative and qualitative parameters of our work during these 10 years is positive. Data are encouraging concerning the possible role of Telemedicine as a tool for domestic assistance of patients affected by chronic diseases, such as CF.

1 INTRODUCTION

Starting out over 40 years ago with demonstrations of hospitals extending care to patients in remote areas, the use of telemedicine has spread rapidly and could now overcome limitations linked with the traditional, restricted and highly expensive in-patient treatment of many chronic pathologies (Castelnuovo, Zoppis, Santoro, Ceccarini, Pietrabissa, Manzoni, Corti, Borrello, Giusti, & Cattivelli 2015, Santoro, Castelnuovo, Zoppis, Mauri, & Sicurello 2015).

In Cystic Fibrosis (CF), the natural history is characterized by recurrent episodes of respiratory infection that causes a progressive pulmonary damage, with decay of long-term lung function leading to death (Flume 2007). Spirometry shows over time in CF patients a reduction in FEV1 (Forced Expiratory Volume in the first second), around 2% of the expected FEV1 value every year, and then a reduction in FVC (Forced Vital Capacity) (Davis, Byard, & Konstan 1997). In case of

pulmonary infection, an early antibiotic treatment helps to prevent more serious complications and consequently to limit the pulmonary damage in the long term. Early intervention also allows us to use advantageously less invasive antibiotic therapies, even using the oral route (Que, Cullinan, & Geddes 2006).

Since 2001, in the CF Center of the Pediatric Hospital Bambino Gesù in Rome, we use Telehomecare (THC) in the follow-up of our patients at home. The first results of this work have been encouraging. We found a statistically significant reduction in hospital admissions and a tendency over time towards a better stability of the respiratory function (Bella & Murgia 2009). In the present study, we examine the data related to monitoring activities on our CF patients followed at home for a period of 4.5 years, in order to investigate the evolution of the clinical trend. The study has the potential to be of great benefit to clinicians, as the effectiveness of Telemedicine in CF population has not previously been reported.

2 MATERIALS AND METHODS

This is a case feasibility study on using Telehomecare in follow-up of CF. We performed an open label trial in a population of CF patients followed in our reference center from 2011 to 2014. Patients were eligible if they have completed the follow-up by THC for the whole period. The intervention study consisted in administering THC in adjunct to standard therapy.

We enrolled a control group among patients seen on the same period, matching for respiratory function, bacterial colonization, sex, age, and complications. The main outcome measure considered in the study was FEV1 values over time. We followed and treated patients included in THC program with the usual protocols, similar to those who do not practice (Bethesda 1997). A clinical diagnosis of CF was given in all subjects, confirmed by study of CFTR (Cystic Fibrosis Transmembrane Conductance Regulator) gene and by sweat test.

We used Spirotek™ instrumentation from MIR (Medical International Research, via del Maggolino, 125, 00155 Roma), which provides and transmits remotely data from spirometry and overnight pulse oximetry. The work flow was as follows.

At home, data are recorded on intervals scheduled with the CF center's physicians, depending on the patient's clinical situation, on average twice a week. Patient may anyway decide autonomously to transmit even without notifying before. Patients perform at home the registration of oxygen saturation and heart rate by night.

In the morning, after chest physiotherapy and mucus drainage, they perform a spirometry, after answering a simple questionnaire regarding some pulmonary symptoms. Data are transmitted by e-mail to the Center. Healthcare professionals trained in telemonitoring download data in hospital once a day using a dedicated application software running on a conventional Personal Computer connected by the hospital intranet to the Net and store data in a local database (Murgia, Cilli, & Renzetti 2011).

As intervention criteria, we considered acute reductions of FEV1 (>10% compared to previous value recorded in stable clinical conditions) (Ramsey & Farrell 1992). For the nocturnal pulse oximetry, we considered significant a fall below 90% of the maximum value of oxygen hemoglobin saturation (SaO2), a reduction of mean SaO2 and an increase over 5% of T90 (percentage of the detection time spent below 90%). Every patient is called back by phone to recall anamnesis data and to share the results. Anamnesis data and graphs obtained are discussed in a briefing between CF Center healthcare professionals for an overall evaluation and to decide on any therapeutic action.

Patients showing significant data are invited to transmit soon further tests. In some cases, if suggested by anamnesis or by data collected, antibiotic home therapy is prescribed based on the last sputum culture. In other cases, patients are invited to the CF Centre for a clinical evaluation, to perform further testing, or to be admitted. In any case, the next data transmission is scheduled. Since February 2010, we started keeping an electronic register, in spread sheet format. For each transmission, the main parameters and the measures are recorded. A monthly statement of assets and the calculation of the average percentage of Adherence to the recommended frequency of transmissions (defined as the ratio transmissions/total patient days) is automatically performed.

3 RESULTS

The current analysis was related to the activity carried out in the period from February, 2010 to June, 2014. Specifically, the target of our study was to check the clinical trend for the THC patients compared to the control group during the last 4 years of follow up¹.

While a direct comparison between case and control group does not provide significant differences (i.e. an independent-samples t-test does not give significant differences between the z score values of the FEV1 for THC patients and controls: $t(27) = 0.73$, $p = 0.47$ in 2011, $t(27) = 0.18$, $p = 0.86$ in 2012, $t(27) = 0.30$, $p = 0.77$ in 2013 and $t(27) = 0.31$, $p = 0.76$ in 2014, see Fig. 1), the results obtained studying the evolution of the scores over time, measured as the difference between the z-score values of the average annual FEV1 and those values (z-scores) observed during the starting reference year (i.e. 2011), returned a positive outcome. More specifically, given the z-score mean value of $\delta_i = k_i - FEV_{i(y)}$ for each subject i , with k_i the base FEV1 value for i at the reference time (i.e. 2011) and y the year of observation (ranging from 2012 to 2014), an independent t-test reported significant differences, with mean values in Fig. 3, for THCs and controls scores (Fig. 4) with the following statistics.

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sign.	t	df	Sign. (2-tailed)	Mean Difference	Std. error difference
Zscore(fev2011)	Eq var	.018	.893	-.731	27	.471	-.27380537	.37474422
	Eq var not assumed			-.734	26.959	.470	-.27380537	.37327051
Zscore(fev2012)	Eq var	.007	.935	-.176	27	.861	-.06668662	.37821313
	Eq var not assumed			-.177	26.954	.861	-.06668662	.37649135
Zscore(fev2013)	Eq var	.131	.720	-.303	27	.766	-.11374811	.37779733
	Eq var not assumed			-.302	26.944	.765	-.11374811	.37622324
Zscore(fev2014)	Eq var	.872	.359	-.307	27	.761	-.11582097	.37777373
	Eq var not assumed			-.308	26.796	.760	-.11582097	.37564525

Figure 1. THC vs control group—years 2011–2014.

¹For the first 5 years, please refer to (Bella & Murgia 2009).

- Year 2011: $t(18.52) = -2.1, p = 0.046$, equal variance not assumed at significance level $\alpha = 0.05$ with Levene's Test;
- Year 2012: $t(16.8) = -2.7, p = 0.014$, equal variance not assumed at significance level $\alpha = 0.05$ with Levene's Test;
- Year 2013: $t(27) = -2.1, p = 0.048$, equal variance assumed at significance level $\alpha = 0.05$ with Levene's Test.

As reported in Fig. 3 the increment of the average scores for THC patients during the observation period provides an evidence to support the use of the THC treatment during the follow-up of the CF disease.

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sign.	t	df	Sig. (2-tailed)	Mean Diff.	Std. error
z-score(Fev2011-Fev2012)	Eq var assumed	12.098	.002	-2.082	27	.047	-.73135783	.35128243
	Eq var not assumed			-2.137	18.520	.046	-.73135783	.34223988
z-score(Fev2011-Fev2013)	Eq var assumed	10.256	.003	-2.662	27	.013	-.89645934	.33681665
	Eq var not assumed			-2.741	16.797	.014	-.89645934	.32700813
z-score(Fev2011-Fev2014)	Eq var assumed	3.249	.083	-2.067	27	.048	-.72695947	.35162041
	Eq var not assumed			-2.089	25.792	.047	-.72695947	.34805296

Figure 2. THC vs control group—tests evaluate the differences of the z score between the reference year (i.e. year 2011) and the period 2012–2014.

	Group	N	Mean	Std. Dev	Std err mean
z-score(Fev2011-Fev2012)	0	15	-.3530693	1.22679446	.31675697
	1	14	.3782885	.48487549	.12958843
z-score(Fev2011-Fev2013)	0	15	-.4327735	1.20678825	.31159139
	1	14	.4636859	.37125693	.09922259
z-score(Fev2011-Fev2014)	0	15	-.3509460	1.06708321	.27551970
	1	14	.3760135	.79574907	.21267289

Figure 3. Group statistics—years: 2011–2014.

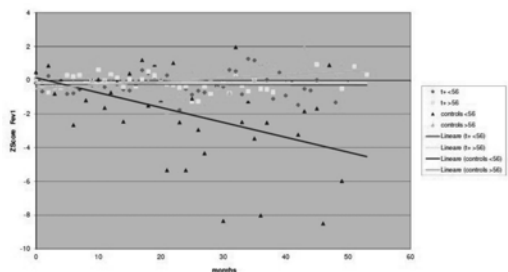


Figure 4. Clinical trial—first 5 years of follow-up.

Moreover, here we also report the trend for the FEV1 monthly means in THC patients and controls during the first follow-up period (5 years) expressed as z-score (Fig. 4). As shown in (Bella & Murgia 2009) the THC-treated group was characterized by higher stability of FEV1 values according to lower SD values of FEV1. This trend might suggest a situation of higher stability of the FEV1 and, therefore, of the respiratory function, in THC treated subjects.

4 DISCUSSION

The results of our study after 10 years of follow-up in Telemedicine showed a significant lesser decline in lung function than those in follow-up with the traditional method. The activity data show an increase overtime of transmissions (Bella & Murgia 2009), despite the number of patients followed with Telemedicine has remained virtually constant. In our experience, one of the critical aspects in the follow-up of chronic patients is a poor adherence to therapy. We highlighted in time a significant increase in adherence to telemonitoring. Patients have accepted the home telemonitoring, intended as innovation in the follow-up, positively. This is evident from the increase in daily telephone responses (as if patient would expect to be contacted by the Center). Pending that institutions are “noticing” the usefulness of telemonitoring, it would be appropriate that systems and procedures are designed and validated by experienced and qualified staff. In our experience, gained over a relatively long period, Telemedicine is a method certainly useful in the follow-up of chronic disease as CF, because it allows:

- A better quality of life
- A lesser deterioration of lung function, with consequent lesser need, in the long term, of invasive therapies.
- A radical change of the motivations of the accesses to the hospital, which become more rational and less demanding both for the patient and for the staff attending to the assistance.

In conclusion, while reliable results on the long term effectiveness of the use of THC in the treatment of CF patients are still lacking, the time has come to obtain data through a multi-center collaboration study, also in order to standardize the international Telemedicine protocols.

REFERENCES

Bella, S. & F. Murgia (2009). Five years of telemedicine in cystic fibrosis disease. *Clini. Ter.* 160(6), 457–460.

- Bethesda, M. (1997). *Cystic Fibrosis Foundation*.
- Castelnuovo, G., I. Zoppis, E. Santoro, M. Ceccarini, G. Pietrabissa, G. Manzoni, S. Corti, M. Borrello, E.M. Giusti, & R. Cattivelli (2015). Managing chronic pathologies with a stepped mhealth-based approach in clinical psychology and medicine. *Frontiers in psychology* 6.
- Davis, P., P. Byard, & M. Konstan (1997). Identifying treatments that halt progression of pulmonary disease in cystic fibrosis. *Pediatr Res* 41(2), 161–5.
- Flume, A. (2007). Cystic fibrosis pulmonary guidelines: chronic medication for maintenance of lung health. *Am J Respir Crit Care* 176, 957–69.
- Murgia, F., M. Cilli, & E. Renzetti (2011). Remote telematic control in cystic fibrosis. *Clin Ter* 162(4), e121–124.
- Que, C., P. Cullinan, & D. Geddes (2006). Improving rate of decline of fev1 in young adults with cystic fibrosis. *Thorax* 61, 155–7.
- Ramsey, B. & P. Farrell (1992). Nutritional assessment and management in cystic fibrosis: a consensus report the consensus committee. *Am J Clin Nutr* 55(1), 108–16.
- Santoro, E., G. Castelnuovo, I. Zoppis, G. Mauri, & F. Sicurello (2015). Social media and mobile applications in chronic disease prevention and management. *Frontiers in psychology* 6.

DIABESITY: Design of mHealth integrated solutions for empowering diabetic and obese citizens in self-monitoring and self-management using mobile devices, apps, social media and web-based technologies

I. Zoppis, G. Mauri & F. Sicurello

Department of Computer Science, University of “Milano-Bicocca”, Milano, Italy

E. Santoro

Laboratory of Medical Informatics, Department of Epidemiology, IRCCS, Mario Negri, Milano, Italy

G. Castelnuovo

Department of Psychology, University “Cattolical del Sacro Cuore”, Milano, Italy

ABSTRACT: Obesity is one of the most important medical and public health problems of our time: it increases the risk of many health complications such as type 2 diabetes, needs long-lasting treatment for effective results and involves high public and private costs. Therefore, it is imperative that enduring and low-cost clinical programs are developed and evaluated.

As reported in several studies, ICT may be valid alternatives to reduce costs and improve adherence to prescribed treatment. Nevertheless few studies have tested a long-term intervention addressed to the behavior change and to measure the weight loss of obese subjects. For this reason, we developed the DIABESITY study, the design of a mHealth integrated platform for empowering diabetic and obese citizens in self-monitoring, and self-management through the use of mobile devices, monitors and treatment protocols. In this paper we focus on the following three important aspects of DIABESITY.

i) Dietary mHealth tools for home-patients; ii) Application and analysis of psychological factors and processes which mediate change of behavior and affect initiation and maintenance phases; iii) Employment of social networks for patients and clinicians.

Currently, this study involves 14 international partners chosen amongst hospitals, universities and ICT companies which will strictly collaborate by contributing with their own specific skills.

The effectiveness of DIABESITY compared with usual care (hospital-based treatment) will be provided in a randomized controlled trial with a 24-month follow-up. In particular, here we report primary and secondary clinical outcomes with the basic statistical procedures which will be used for this evaluation.

1 INTRODUCTION

Obesity and low level of physical activity are well known important public health problems that increase the risk of many complications such as hypertension, hypercholesterolemia, coronary artery disease, and type 2 diabetes (Consortium InterAct 2012). World-wide, nearly 250 million people have diabetes, and the number of people affected by this disease is increasing rapidly. According to the World Health Organization the prevalence of diabetes for all age-groups is estimated to be 4.4% in 2030. The total number of people with diabetes is projected to rise to 366 million in 2030. Serious health consequences, in turn, weigh heavily on public health care costs. According to the IDF Diabetes Atlas, (fifth edition) the estimated global health care expenditures to treat diabetes and prevent compli-

cations totaled at least US Dollars (USD) 465 billion in 2011. By 2030, this number is projected to exceed some USD 595 billion.

Cost-effective approaches that can reach broad populations of obese people are thus needed and have to be evaluated overall with regard to compliance and healthy behavior maintenance in the long term. In fact, promising methods for granting continuity of care to wide populations of patients at low costs are provided by telemedicine and its more specific branches called “e-therapy”, “telecare” and “e-health” (Pagliari, Sloan, Gregor, Sullivan, Detmer, Kahan, Oortwijn, & S. 2005, Eysenbach 2001).

As already indicated in several studies (A.D. & M. 1999, Rice 2005), these technologies may be valid alternatives to reduce expensive and time-consuming clinical visits and to improve adherence to prescribed treatment through extensive monitoring

and support. In particular, social media can directly support disease management by creating online spaces where patients can interact with clinicians and share experiences with other patients (Coeira 2013, Li J. S. 2013, Santoro, Castelnuovo, Zoppis, Mauri, & Sicurello 2015). Cancer patients use Twitter to discuss treatments and provide psychological support (Sugawara, Narimatsu, Hozawa, Shao, Otani, & Fukao 2012) and online engagement seems to correlate with lower levels of self-reported stress and depression (Beaudoin & Tao 2008).

Nevertheless much more work remains to be carried out in order to confirm these findings. For example, few studies have investigated the effects of an ICT-based program by measuring the weight loss of obese patients with established type 2 diabetes. Furthermore, to the best of our knowledge, few programs have investigated a comprehensive long-term stepped down intervention specifically addressed to the behavior change (Castelnuovo, Manzoni, Cuzziol, Cesa, Tuzzi, Villa, Liuzzi, Petroni, & Molinari 2010, Castelnuovo, Manzoni, Corti, Cuzziol, Villa, & Molinari 2011).

According to Ryan (Ryan, Patrick, Deci, & Williams 2008) there are many approaches that have proved to be effective in initiating change, from external pressure and control to the positive use of incentives or rewards, but the essential ingredients to maintenance are still missing. For these reasons, we developed DIABESITY, a detailed study of mHealth

(Mobile Health) services provided by an innovative platform where different technologies for the disease and the intervention management are integrated.

In this paper, we describe how this platform can help patients in maintaining lifestyle behavior changes, ensuring functional patient empowerment and engagement. With this aim, we report the following key points integrated within DIABESITY:

- i. Dietary mHealth services for home-patients (section 2.2);
- ii. Psychological and behavioral questionnaires and indexes for change of behavior quantification (section 2.3);
- iii. Social networks services prospective for patients and clinicians (section 2.4).

Finally, in section 2.5, we report the primary and secondary clinical outcomes with the basic statistical procedures which will be used for analyzing the effectiveness of this study.

2 MATERIAL AND METHODS

2.1 Design

DIABESITY involves 14 international partners chosen amongst hospitals, universities and ICT companies, specialized in software development, management, medical and clinical expertise. The

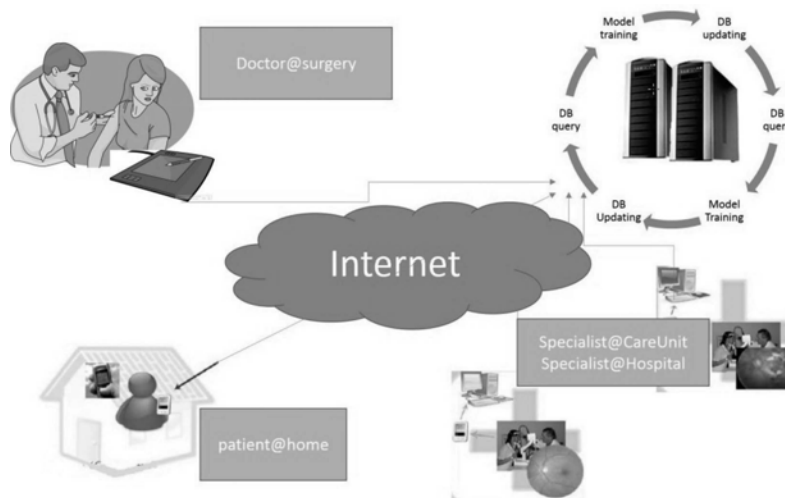


Figure 1. System Architecture. In DIABESITY we can distinguish network elements, monitors and telecommunication interfaces, specific devices (e.g. wearable) and a central webservice with hardware for hosting databases and mining procedures. The platform is based on the idea to implement an innovative web-based system on mHealth (Mobile-Health) technologies for the disease management and intervention. Web-based services will provide interoperability between patients (overweight/obese subject and diabetics) and specialists as well as easy health professionals access and sharing of biomedical knowledge.

Consortium has an established international reputation in the field of diabetes, and a solid track in participation to European projects. Figure 1 illustrates the architecture of DIABESITY.

The system should be developed as integrable platform that should fit into existing diabetes health system functions and complement the health system goals of health service provision for overweight/obese people and patients with type 2 diabetes. Moreover DIABESITY will strictly follow current relevant standards, guidelines, and best practices, in particular those concerning interoperability, minimum patient summary dataset to be shared across borders, standard on user safety (currently draft standard IEC 82304-1), appcertification programs (e.g. NHS in the UK), apps as medical device (directive 93/42/EC under review) or in vitro diagnostic medical device (directive 98/79/EC under review), compliance with personal data protection rules, etc.

The theoretical approach of DIABESITY is trying to “move the healthcare where it really needs” using advanced tools to ensure for patients the continuity of treatment at home, using desktop computers and mobile phones connected with clinicians (psychologists, endocrinologists, dieticians) that have already attended the patients inside hospitals or during brief standard out-patient care.

Here we do not focus on the technologies represented in Figure 1, nor on how that architecture will be integrated; rather in next paragraphs we specify which functionality of DIABESITY will be applied in order for patients to maintain significant lifestyle behavior changes, improve health outcomes, and ensure functional empowerment and engagement.

2.2 Dietary mHealth tools for home-patients

Most of the studies showed the pros and cons of the existing mHealth applications. Through the employment of statistical and Data Mining (DM) techniques (build on advanced relational and heterogeneous information-based models to cluster group of similar instances, e.g. (Antoniotti, Carreras, Fari-naccio, Mauri, Merico, & Zoppis 2010, Cava, Zoppis, Gariboldi, Castiglioni, Mauri, & Antoniotti 2014, Zoppis, Merico, Antoniotti, Mishra, & Mauri 2007)) the DIABESITY technology will address the following not yet solved issues.

- Integration of data from sensors and manual input from users;
- Evaluation of temporal development through trend analysis—a tool for user motivation and personalized recommendations;
- Advanced features on user interface such as visualization of self-monitoring data;

- Use of the information collected by mobile applications to monitor progression of the disease and/or treatment and its impact on the patient’s lifestyle.

Specifically next paragraphs report the main apps of DIABESITY.

2.2.1 Diet caloric restriction

This module is aiming to promote revolutionary principle approach of eating low calories density food, i.e. “eat more food while eating fewer calories and feeling the same degree of satiety” to maintain normal weight. The following functions should be provided.

- Tables and graphs stating that subjects have a confidence to undergo a specific dietary prescription, given his/her current dietary or behavioral condition. In this case, the consumer will get a combination of likely successful dietary treatments from a set of patients feature characteristics;
- Makes it easy to find a partner doctor in one of National’s healthcare networks around-the-clock, anywhere and access to that doctor for advice;
- Enables patients to quickly book appointments at a Health Diabetes, Nutrition and Metabolic Diseases Center;
- The app users will be able to have secure access to their health data and logs via the app.

2.2.2 Diet composition

In this case, the app focuses on improvement in dietary patterns national diet, in increase vegetables and implementing existing knowledge in diet and how many changes will be needed for diabetes prevention (intervention pattern in fat, sugar intake and micronutrients intake). A healthy diet (e.g. consume more fibre and less saturated fat and dietary cholesterol) reduces risk of CVD possibly by modification of endothelial dysfunction and low-grade inflammation (play central role in CVD). The app offers the following functions.

- Tables and reports stating that subjects need improvement in dietary patterns and intervention pattern in fat, sugar intake and micronutrients in take. This report suggest confidence to undergo a specific dietary improvement, given his/her current dietary or behavioral condition. In this case, the consumer will get a combination of likely successful dietary improvement from asset of patients feature characteristics;
- Makes it easy to find a partner dietician (a registered dietitian) in one of National’s Health

Diabetes, Nutrition and Metabolic Diseases Center networks around-the-clock, anywhere and access to that specialist for advice;

- Dieticians can make a personalized meal plan or can recommend standardized food portion size (fibre, carbohydrate, fat, cholesterol, protein, alcohol, total energy) and can send images of the dietary patterns simply via the app;
- The app users will be able to create a personal patient record with standardized food portion size (or use already existing one in the system) and have secure access to their health data via the app.

2.2.3 Home exercise postnatal factors

Individuals could benefit from the exercises provided and optimized by the knowledge discovery (or Data Mining) algorithms, as well as logs and exercise sessions. Consumers, may choose pre-designed programs based on preferences and goals as well as from appropriate personalized and guided workouts, performance monitoring, action-directed instructions. The app offers the following functions:

- Tables and graphs stating that subject has a confidence to undergo a specific behavioral prescription, given his/her current clinical, dietary or behavioural condition. In this case, the consumer will get a combination of likely successful treatments from a set of patients characteristics;
- Makes it easy to find a partner obstetrical doctor or physical trainer in one of National's healthcare networks around-the-clock, anywhere and access to that doctor for advice;
- Enables patients to quickly book appointments at a Health Diabetes, Nutrition and Metabolic Diseases Center;
- The app users will be able to have secure access to their health data and logs via the app;
- Connectivity to wearable devices in noninvasive medical monitoring (e.g. glucose monitors, BP monitors).

2.3 Psychological and behavioral questionnaires

As suggested by Katan (Katan 2009) with regard to dieting, cognition and feelings have a huge impact on behavior and may thus strength as well as disrupt adherence to treatment and compliance with clinical prescriptions. Indeed, psychological factors and processes mediate every behavior change and differently affect both the initiation and maintenance phases (Rothman 2000).

DIABESITY program is designed to explore and measure such factors with the further important objective to define the type of patients that benefit the most from such an intervention. For this, the following questionnaires and indices will be employed.

- *The Self-Report Habit Index (SRHI)*—Italian translation.

The SRHI is a measure of the development and strength of habits. It has a stem “the behavior is something that...” followed by 12 items such as “I do without thinking”. The SHRI has high internal consistency ($\alpha > 0.9$), high test-retest reliability ($r = 0.91$)¹, high convergent and discriminative validity (Verplanken & Orbell 2003). In our study, behaviors are: “eating in accordance with the prescribed diet” and “undertaking regular physical activity”.

- *Weight Efficacy Life Style Questionnaire (WELSQ)*—Italian version.

The WELSQ is composed of 20 items that measure the confidence of the subjects about being able to successfully resist the desire to eat. The questionnaire was used to predict both weight loss and weight loss maintenance across a range of ages in men and women (Castelnuovo, Gaggioli, Mantovani, & Riva 2003).

- *Body Uneasiness Test (BUT)*—Italian version.

The BUT is a self-report inventory that measures body uneasiness by a global severity index and five sub-scales: Weight Phobia, Body Image Concerns, Avoidance, Compulsive Self-Monitoring, Depersonalization (Cuzzolaro, Vetrone, Marano, & Battacchi 1999).

- *Binge Eating Scale (BES)*—Italian version.

The BES is a short self-report questionnaire which measures severity of binge eating (Gormally, Black, Daston, & Rardin 1982).

The Italian version of the instrument (Di Bernardo, E., Ricca, Mannucci, Moretti, Cabras, & Rotella 1997) consists of 16 items, which are composed by three or four sentences about the severity of binge eating. Cut-off score for mild binge eating symptoms is 17; scores between 18–26 indicate moderate binge eating symptoms and scores over 27 can be associated with a severe binge eating disturbance.

- *Eating Disorder Inventory (EDI-2)*—Italian version.

The EDI-2 is a widely used, standardized, self-report measure of psychological symptoms commonly associated with an or exianervosa, bulimia nervosa and other eating disorders. The EDI-2 does not yield a specific diagnosis of eating disorder. It is aimed at the measurement of psychological traits or symptom clusters presumed to have relevance to understanding and treatment of eating disorders. The EDI-2 consists of 11 subscales derived from 91 items. Three of the subscales were designed to assess attitudes and

¹Reliability value (r) refers to the reproducibility of values of a variable when the same object is measured more than once (Hopkins 2000).

behaviors concerning eating, weight and shape (Drive for Thinness, Bulimia, Body Dissatisfaction) and the remaining eight ones tapped more general constructs or psychological traits clinically relevant to eating disorders (Ineffectiveness, Perfection, Interpersonal Distrust, Interoceptive Awareness, Maturity Fears, Asceticism, Impulse Regulation and Social Insecurity)(Conti 2002, Garner 1991).

- *Symptom Check List (SCL-90)*—Italian version.

The SCL-90 is a brief, multidimensional self-report inventory designed to screen for a broad range of psychological problems and psychopathological symptoms. It consists of 9 symptom scales (Somatization, Obsessive-Compulsive, Interpersonal Sensitivity, Depression, Anxiety, Hostility, Phobic Anxiety, Paranoid Ideation and Psychoticism) and 3 global indices (Derogatis, Lipman, & Covi 1973).

- *Impact of Weight on Quality of Life-Lite (IWQOL-Lite)*—Italian version.

IWQOL-Lite is the short version of the original IWQOL and is composed by 31 items. The questionnaire is self-report and consists of 5 scales assessing the impact of weight on QoL-related factors such as Physical Functioning, Self-Esteem, Sexual Life, Public Distress and Work. IWQOL-Lite has shown high internal consistency and high test-retest reliability (Kolotkin, Crosby, Kosloski, & Williams 2001, Kolotkin, Crosby, & Williams 2002).

- *The Outcome Questionnaire (OQ 45.2)*—Italian translation.

The OQ 45.2 is a self-report questionnaire developed by Michael Lambert in 1996 (Lambert, N.B., Umphress, Lunnen, Okiishi, Burlingame, & Reisinger 1996). The OQ 45 items version is a measure of outcome and it is designed in order to collect repeated measures of patient progress during therapy and after its conclusion.

This instrument is one of the most used in psychotherapy research in the U.S. (Hatfield & Ogles 2004). The OQ 45.2 is composed by 45 items that form 3 scales: Symptom Distress (SD), Interpersonal Relations (IR) eSocial Role (SR), and a Global Index.

2.4 Social networks

Social networks can directly support disease management. DIABESITY is designed to integrate two main communities.

1. Patient community—where patients can post and share non-clinical data (information on improvement to diet and physical activity, opinions, suggestions, comments and questions).

This community will be moderated by specific health professionals who also will routinely post information (such as educational content, research news, current diabetes fact sheets) to help patients.

2. Scientific and clinical community—Where the following main objectives will be provided:

- Dissemination of the results through article published in high-impact journals, and through international diabetes conferences;
- Publishing technological results of the platform, in this case to influence policymaker and reimbursements.

2.5 Statistical analysis

To date there have been few trials that have tested the efficacy of telemedicine in the field of obesity with type 2 diabetes (or with those at risk of diabetes). In order to evaluate the efficacy of this study we apply the following design.

200 patients will be randomly allocated to the following 2 groups:

1. DIABESITY group: Extensive outpatient telecare through the web-platform and mobile phones;
2. CONTROL group: In-hospital treatment (diet, physical activity, psychological and dietitian counseling) and follow-up assessment.

Patients will be recruited within three months from the start of the trial and will be followed up for one year. Patients and physicians will be trained to use the mHealth components of the DIABESITY platform and the clinical trial data management tools. The clinical trials will be registered in the ClinicalTrials.gov register and in the national registries of each participating/partner country.

Various data will be collected during the trial including anamnestic data, physiological and biometric (sensors) data, glycemic and dietary data, exercise training data and quality of life data. Specifically,

- Primary outcome will be based on comparison of Glycated hemoglobin (HbA1c) and weight values at baseline to HbA1c and weight values at study end (30 months).
- Secondary outcome measures will be given through energy expenditure, glycated hemoglobin, binge eating, self-efficacy in eating and weight control, body satisfaction, healthy habit formation, disordered eating-related behaviors and cognitions, psychopathological symptoms and weight-related quality of life.
- System data (apps and social media) will be used to correlate with changes in primary and secondary outcome measures.

Statistical analysis will be performed on primary and secondary outcomes in order to investigate the DIABESITY efficacy vs traditional care on changes of Glycated hemoglobin (HbA1c) and weight values from baseline to one year of follow-up. Procedures for data anonymization and data integration (semantically linked) with other source of (future) different information will be also considered.

Descriptive statistics (means SD, or median and interquartile ranges, as appropriate) will be used to report the study sample with regard to baseline characteristics. Before selecting the most appropriate statistical tests, assumptions for parametric/non-parametric analysis will be checked.

Repeated-measure ANCOVA will be used in order to evaluate the effects of the intervention when data do not violate the parametric assumptions. The mean differences between intervention and control group with 95% confidence intervals will be calculated. Analysis will be adjusted for possible confounders such as gender and age. Also effect modification will be investigated using interaction terms between intervention group and gender and age, respectively. If data violate parametric assumptions, non-parametric statistical tests will be applied to make reliable inferences.

All data analysis will be performed using Statistical Packages currently applied for data analysis and data mining procedure (SPSS and R).

3 CONCLUSIONS

Intelligent technologies and devices (for example automated analysis, medical tools that can self-monitor and call upon expert/professional help) will play an increasing role in health-care in the near future. Moreover, miniaturization of diagnostic and monitoring tools is likely to be significant, making them available in local or home settings. For these reasons new studies and projects are fundamental to better understand the efficacy of new technologies and the type of patients that benefit the most from such an intervention.

Within DIABESITY we integrate three different solutions for both the management and intervention of obese/overweight subject with diabetes. This integration is designed to promote and measure the weight loss of obese patients and the psychological factors and processes which mediate behavior changes and affect initiation and maintenance phases.

Specifically, DIABESITY explores the long-term intervention addressed to measure the weight loss of obese patients and the psychological factors and processes which mediate behavior changes and affect initiation and maintenance phases.

This study is a combined effort to merge the experience and skills of international partners with the usability, scientific and exploitation needs of the end users. In fact, the basic idea behind DIABESITY program is to “move the healthcare where it really needs”, thus offering important social, political and technological impacts for examples:

- Providing regular feedback, follow-up, and pragmatic advice for the target population.
- Support for healthcare providers to focus on updating and reinforcing knowledge about the future health implications of DIABESITY.
- Significant knowledge progresses related to psychological and medical variables, and processes about efficacy and effectiveness of the use of telecare and social media tools in developed care models, specifically about obesity treatment and monitoring in traditional and technology-based settings.
- Reduced hospitalization rate or hospital attendance in out-patient settings and improved disease management, treatment and rehabilitation at the point of need, through more precise assessment and monitoring of health status.
- Improved quality of life and patients empowerment.
- Improved links and interactions between patients and clinicians facilitating more active participation of patients in care processes (patient engagement and empowerment).

Finally this program will be evaluated in a randomized controlled trial. We will test the effectiveness of this stepped down program enhanced by telecare on weight loss, weight loss maintenance, energy expenditure, glycated hemoglobin, binge eating, self-efficacy in eating and weight control, body satisfaction, healthy habit formation, disordered eating-related behaviors and cognitions, psychopathological symptoms and weight-related quality of life.

The authors declare that there is no conflict of interests regarding the publication of this manuscript.

REFERENCES

- A.D., C. & W.M. (1999). New frontiers in using telemedicine for nutrition intervention. *J Am Diet Assoc.* 99(11), 1442–1443.
- Antonioti, M., M. Carreras, A. Farinaccio, G. Mauri, D. Merico, & I. Zoppis (2010). An application of kernel methods to gene cluster temporal meta-analysis. *Computers & operations research* 37(8), 1361–1368.
- Beaudoin, C. & C. Tao (2008). Modeling the impact of online cancer resources on supporters of cancer patients. *New Media Soc.* 10, 321–44.
- Castelnuovo, G., A. Gaggioli, F. Mantovani, & G. Riva (2003). From psychotherapy to e-therapy: the integra-

- tion of traditional techniques and new communication tools in clinical settings. *Cyberpsychol. Behav.* 6, 375–382.
- Castelnuovo, G., G. Manzoni, P. Cuzziol, G. Cesa, C. Tuzzi, V. Villa, A. Liuzzi, M. Petroni, & E. Molinari (2010). Tecnob: study design of a randomized controlled trial of a multi-disciplinary telecare intervention for obese patients with type 2 diabetes. *BMC Public Health*. 20(1).
- Castelnuovo, G., G. Manzoni, S. Corti, P. Cuzziol, V. Villa, & E. Molinari (2011). Clinical psychology and medicine for the treatment of obesity in out-patient settings: The tecnob project. *Telemedicine Techniques and Applications*.
- Cava, C., I. Zoppis, M. Gariboldi, I. Castiglioni, G. Mauri, & M. Antoniotti (2014). Combined analysis of chromosomal stabilities and gene expression for colon cancer progression inference. *Journal of clinical bioinformatics* 4(1), 1.
- Coeira, E. (2013). Social networks, social media, and social diseases. *BMJ* 346, f3007.
- Conti, L. (2002). *Repertorio delle scale di valutazione in psichiatria*. Firenze, Italy: S.E.E.
- Cuzzolaro, M., G. Vetrone, G. Marano, & M. Battacchi (1999). But: Una nuova scala per la valutazione del disagio relativo all'immagine del corpo in adolescenza. *Psichiatria dell'infanzia edell'adolescenza*.
- Derogatis, L., R. Lipman, & L. Covi (1973). Scl-90: an outpatient psychiatric rating scale—preliminary report. *Psychopharmacol Bull* 9(1), 13–28.
- Di Bernardo, M., B.E., V. Ricca, E. Mannucci, S. Moretti, P. Cabras, & C. Rotella (1997). Validazione della versione italiana della binge eating scale in pazienti obesi. *Minerva Psichiatrica*. 39, 125–130.
- Eysenbach, G. (2001). What is e-health? *J. Med Internet Res*. 3(2), e20.
- Garner, D. (1991). Eating disorder inventory-2. edi-2. *Professional Manual*.
- Gormally, J., S. Black, S. Daston, & D. Rardin (1982). The assessment of binge eating severity among obese persons. *Addictive Behaviors* 7(1), 47–55.
- Hatfield, D. & B. Ogles (2004). The use of outcome measures by psychologists in clinical practice. *Professional Psychology: Research and Practice*. 35(5), 485–491.
- Hopkins, W. (2000). Measures of reliability in sports medicine and science. *Sports Medicine*. 30, 1–15.
- Katan, M. (2009). Weight-loss diets for the prevention and treatment of obesity. *N Engl J Med*. 360(9), 923–925.
- Kolotkin, R., R. Crosby, & G. Williams (2002). Health-related quality of life varies among obese subgroups. *Obes. Res*. 10(8), 748–756.
- Kolotkin, R., R. Crosby, K. Kosloski, & G. Williams (2001). Development of a brief measure to assess quality of life in obesity. *Obes. Res*. 9(2), 102–111.
- Lambert, M., H.N.B., V. Umphress, K. Lunnen, J. Okiishi, G. Burlingame, & C. Reisenger (1996). Administration and scoring manual for the outcome questionnaire (oq45.2). *Stevenson MD: American Professional Credentialing Services LLC*.
- Li J.S., e. a. (2013). Approaches to the prevention and management of childhood obesity: the role of social networks and The use of social media and related electronic technologies: a scientific statement from the American heart association. *Circulation* 1272, 260–7.
- Pagliari, C., D. Sloan, P. Gregor, F. Sullivan, D. Detmer, J. Kahan, W. Oortwijn, & M.S. (2005). What is health (4): a scoping exercise to map the field. *J. Med Internet Res* 7(1), e9.
- Rice, C. (2005). Prevention: the most economical way to manage diabetes. *Nurs Econ* 23(6), 327–329.
- Rothman, A. (2000). Toward a theory-based analysis of behavioral maintenance. *Health Psychol.* 19(1 (Suppl)), 64–69.
- Ryan, R., H. Patrick, E. Deci, & G. Williams (2008). Facilitating health behaviour change and its maintenance: Interventions based on self-determination theory. *The European Health Psychologist*. 10, 2–5.
- Santoro, E., G. Castelnuovo, I. Zoppis, G. Mauri, & F. Sicurello (2015). Social media and mobile applications in chronic disease prevention and management. *Frontiers in psychology* 6.
- Sugawara, Y., H. Narimatsu, A. Hozawa, L. Shao, K. Otani, & A. Fukao (2012). Cancer patients on twitter: a novel patient community on social media. *BMC Res Notes*. 5.
- Verplanken, B. & S. Orbell (2003). Reflections on past behavior: A self-report index of habit strength. *Journal of Applied Social Psychology*. 33, 1313–1330.
- Zoppis, I., D. Merico, M. Antoniotti, B. Mishra, & G. Mauri (2007). Discovering relations among go-annotated clusters by graph kernel methods. Volume 4463 of *Lecture Notes in Computer Science*, pp. 158–169. Springer.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Mining complex networks: A new challenge for supporting diagnostic decisions

I. Zoppis, G. Mauri & F. Sicurello

Department of Computer Science, University of “Milano-Bicocca”, Milano, Italy

E. Santoro

Laboratory of Medical Informatics, Department of Epidemiology, IRCCS, Mario Negri, Milano, Italy

G. Castelnuevo

Psychology Research Laboratory, Istituto Auxologico Italiano IRCCS, Ospedale San Giuseppe, Verbania, Italy

ABSTRACT: New technologies are multiplying at an enormous speed and the produced data is not only massive but also complex. In fact, despite the abundance of tools to capture, process and share information (e.g. data) one cannot broadly assume the standard hypothesis that such data are identically and independently distributed (i.i.d.). As a result, proper handling of data is fundamental in order to convert the available observation in to useful information that leads to knowledge and suitable decision making. In this paper, we focus on network data. That is, we introduce the reader to a theoretical perspective concerning the knowledge mining of huge amount of relational information collected in all the network systems which are ubiquitous in our life. In this context, following a numerical evaluation we show the reader how different kind of information can provide a benefit for a typical machine learning problem i.e. classification. The main issue of our investigation is to provide a case where the accuracy of a classification model benefits when considering the additional information given by both network and dissimilarity features. Moreover, we treat a clinical example that will serve as running case for our analysis.

1 INTRODUCTION

Relationships are undoubtedly central to our lives and new technologies make these relationships to emerge and express their benefits at different levels. Just to mention a few examples, Information and Communication Technology (ICT) have permitted social communities to grow, not only allowing communication among subjects, but even supporting e.g. patient engagement and empowerment (Lamprinos, C., Schmuhl, Demski, Hildebrand, & Plossnig 2014, Castelnuevo 2010, Castelnuevo & Simpson 2011, Castelnuevo, Manzoni, Pietrabissa, Corti, Giusti, Molinari, & Simpson 2014, Santoro, Caldarola, & Villella 2011, Santoro 2013, Santoro & Quintaliani 2013). At a different level, bio-technologies (such as microarrays) not only provided the expression of thousand of genes simultaneously, but even allowed a better understanding of the proper mechanisms which characterize many interactions in different biological systems. But new technologies are also creating, in this context, new challenges from a different perspective: the need to develop methods for managing and mining huge amount of relational information collected in all the network systems

which are ubiquitous in our life. In fact, despite the abundance of tools to capture, process and share information (e.g. data)

– from sensors to computers, from microarray to mobile phones, from a theoretical viewpoint one cannot broadly assume the standard hypothesis that such data are identically and independently distributed (i.i.d.). Instead samples seems to be, more often than not, interconnected each others, thus violating the i.i.d. assumptions.

It is in this context that a recent focus in *machine learning* (Mitchell 1997, Kolaczyk 2014, Witten, Frank, & Hall 2011, Marsland 2011) has been to extend traditional problems to complex interaction systems and more generally to networks (Getoor & Taskar 2007, Bansal, Blum, & Chawla 2002). In this case traditional algorithms not only may benefit of information provided by links between instances but even could fail to give an accurate inference (Zoppis, Merico, Antoniotti, Mishra, & Mauri 2007, Zoppis & Mauri 2008).

In this paper, we introduce the reader to this research providing a real-case study to better understand how to mining knowledge in complex

networks. In particular we show the reader how a particular type of information can provide a benefit for a typical machine learning problem i.e. classification. With this goal we give an example where the accuracy of a classification model improve when considering additional information provided by both network and dissimilarity features (Pekalska & Duin 2005). Specifically, the following points are developed.

1. We describe the Breast Cancer Data-set by giving a network representation (BD network) finalized to apply two-class classification problems;
2. We describe network features for BD;
3. We introduce dissimilarity features extending both the set of network and BD features;
4. We apply a collective classification algorithm (see for example (Indyk, Kajdanowicz, Kazienko, & Plamowski 2012)) for the BD network.

Our evaluation is empirical. First we observe the accuracy of a standard inference method (in this case Partial Least Square as a base and common tool on which we set up our numerical experiment and comparisons) to forecast the case/control membership group of classification instances, when patients (i.e. instances) are represented through the set of available clinical attributes. We call this approach standard (shortly SA) since this reflects the typical way of representing i.i.d. subjects. Then we check whether the inference accuracy improves when explicitly exploiting the classification (i.e. PLS-based collective classification) by adding the information provided through the network and dissimilarity features. We organize the paper as follows.

In section 2.1 we introduce the case-study and the required definitions. In particular, sections 2.3 and 2.3 focus on a typical machine learning problem for networks, i.e. the *classification* problem. The main notation and nomenclature is given in 2.4. In section 2.5 we introduce the dissimilarity features. In section 3 we report the accuracy of the inference systems for our case-study. We conclude the paper in section 4 discussing the obtained results.

2 MATERIAL AND METHODS

While most of the material in this section is mathematical in nature, specialized terminology and symbols should be avoided as much as possible. Our purpose here is to describe, without entering into technical details, an important issue of the machine learning community: the classification problem. We start by describing the Breast data set that will serve as running (real) case-study to provide accurate diagnostic decisions.

2.1 The Breast dataset

The case-study discussed in this paper is the Wisconsin breast cancer data set. Data were originally obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg, who periodically reported his clinical cases (see e.g. (Mangasarian & Wolberg 1990)). The data are also available from the UCI Repository of machine learning databases at <https://archive.ics.uci.edu/ml/index.html>, and are part of the R package *mlbench* (Newman, Hettich, Blake, & Merz 1998). The whole dataset consists of 699 samples; 16 samples, however, have been removed because of missing values. A number of covariates are stored along with data. In particular it is reported the value of nine cytological characteristics graded on a 1 to 10 scale at the time of sample collection, with 1 being the closest to normal tissue and 10 the most anaplastic. In detail, the following attributes are considered clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli, mitoses. The aim of the data analysis is to classify each instance as benign or malignant using these covariates.

2.2 Information for classification problems in complex networks

The goal of machine learning is to discover rules (*knowledge*) using previous experience. When it comes to work with data this issue can be translated in the task of identifying patterns (functional dependencies, relations etc.) from past observations (*experience*), and then to perform useful inference using those patterns that have been *learned*. Such inference typically takes the form of *classification*, which is probably the oldest and most studied problem of all the knowledge discovery tasks. *Classification* is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a set of data containing observations (or instances) whose category membership is known (training set).

A typical classification within healthcare is the determination of whether a new subject can be discriminated from *case* (diseased) patients or classified as *control* according to the previous set of sampled attribute values (see for example (Cava, Zoppis, Gariboldi, Castiglioni, Mauri, & Antoniotti 2013)). Considering our case-study, we may be interested to the problem of predicting if a new subject may be considered as benign or malignant using the covariates reported in BD. The forecasting process is provided by a *classification algorithm* which takes a set of *labeled examples* as input (i.e. patient records with its associated

labeled groups), and build a model (i.e. *training phase*) to assign label sat previously un seen unlabeled examples (i.e. *test set*)¹.

The generally tendency in machine learning has been towards adoption of the i.i.d. assumption, when the inference on the case/control membership group of a new subject is based on the values of personal attributes without taking into account any underlying relationships connecting subjects. In this case one generally refers to *Independent Classification* (IC) since traditional classification algorithms would ignore correlations represented by these interconnections.

When links among instances (patients) are considered, either by expressing some natural or social properties (e.g. environmental, parental or behavioral if available) or explicitly measuring some similarity (or dependencies) between personal attributes (e.g. concerning similar physical training, lifestyle, dietary information or even correlation between clinical traits from different subjects) we generally refer to *Collective Classification* (CC) problem. Thus, the class label of a particular instance depends on the class labels and sometimes even attributes of other related instances and not just on its own set of attribute values. In this case traditional algorithms not only may benefit of information provided by links (Zoppis, Merico, Antoniotti, Mishra, & Mauri 2007, Zoppis & Mauri 2008) but even could fail to infer accurately the discrimination task.

Following our case-study, we should predict the membership group of a particular patient, first by measuring the correlation between patients' covariates, and then by applying a specific CC algorithms (Indyk, Kajdanowicz, Kazienko, & Plamowski 2012, Jensen, Neville, & Gallagher 2004, Macskassy & Provost 2007), which in turn consider (i.e. input) as connected those patients whose correlations are higher than some specific significant threshold. In other words, when the task is to assign each new subject with a label that best indicates if he will be considered as *case* while following standard classification schemes that labeling can be obtained by applying inference only to the patient's record, more accurately one can assume that labels are interconnected (e.g. introducing correlations) and then apply inference mechanisms (algorithms) to the collective systems (i.e. collective classification).

Note that, as mentioned above, such interconnections occur naturally in many data from a

variety of applications such as bibliographic data, email networks, and social networks. In this case, instances are naturally linked each other (e.g. by construction web pages are interconnected through links, Face-book users may interact by listing each other as friends, sending a message, posting on a wall, engaging chat, and soon.).

In other cases links are explicitly introduced. For instance, in our study, for each pair of patients we first obtain the correlation between the irrecord values (i.e. correlation between patient covariates) then we interconnect patients when correlations were higher than a significant threshold².

2.3 The CC paradigm

Collective classification refers to the classification of linked entities in a networks. It is based on the homophile hypothesis, in which linked entities have a tendency to belong to the similar class. In this paper we apply a simple, efficient and widely used method to solve this problem: the Iterative Classification Algorithm (ICA). It is representative of a family of methods for which inference proceeds as an iterative process: at each step, entities of the network are classified according to the current predicted labels of their neighbors. As reported above, here the target is not to detail the collective classification paradigm or the specific methods applied (specifically, the ICA). Instead we applied the Collective classification as a black box procedure, together with PLS (as inference model) with the purpose to take into account of the information provided through the network attributes.

2.4 Some theoretical issues

In this section we introduce the main notation and nomenclature. This will allow us to discuss the numerical results in a consistent manner.

2.4.1 Graphs

For mathematical purposes, a network is most commonly represented in a formal manner using abstract objects called *graphs* (see for example (Bolloba's 1998, Gross & Yellen 2005, Diestel 1997)). Graphs represent networks through a set of nodes (also called vertexes) and edges (also called links). Nodes may possess characteristics which are of interest (such as protein structure or function). Edges may possess different weights, depending on for example, the strength of the interaction or its reliability. Mathematically, we refer to a graph as $G = (V, E)$, where V is these to f nodes and E is these of edges. We use the notation $|S|$ to denote the number of elements in the set S . Then $|V|$ is the

¹Please note that, here we do not deal with classification algorithms. Instead our target is to focus on how to extract suitable in formation from complex networks for providing inference (supporting decisions) in classification problems. A description of classification algorithms can be found e.g. in (Mitchell 1997).

²Statistical significant level.

number of nodes, and $|E|$ is the number of edges in the graph G . If u and v are two nodes and there is an edge from u to v , then we write that $(u, v) \in E$, and we say that v is a *neighbor* of u ; and u and v are adjacent. Edges may be directed or undirected; here we shall mainly deal with undirected edges. While some authors assume that, in contrast to graphs, networks are connected, here we make no such assumption and use the terms graph and network interchangeably.

Characterization of network through numerical summaries (in the following were referred to the quantities as network statistics, network properties or even network attributes) is an important aspect of descriptive analysis for networks. However, unlike in an introductory statistics course, where the types of summaries typically are just simple measures of center (e.g., mean, median, etc.) and dispersion (e.g., standard deviation, range, etc.), summary measures for networks seek to capture characteristic properties of a graph. Although large networks are typically high dimensional and complex objects, many of their important properties can be captured by calculating relatively simple summary statistics. In next section we shall deal with such notions mainly focusing on the graph properties applied for our case study.

2.4.2 Network statistics

Many questions that might be asked about a vertex in a network essentially seek to understand its importance. For instance we might be interested in the following questions. How authoritative does a particular page in the World-Wide-Web seem to be considered? The deletion of which genes in a gene regulatory network is likely to be lethal to the corresponding organism? How critical is a given router in an Internet network to the flow of traffic?

Measures of centrality are designed to quantify such notions of importance and thereby facilitate the answering of such questions. Here we say that a vertex is important (or central) if it is close to many other vertices. There are many number of different centrality measures that have been proposed in literature but probably the most applied is called *vertex degree*. The degree $d(v)$ of a vertex v , in a network $G = (V, E)$, counts the number of edges in E incident upon v .

Given G , define $f(d)$ to be the fraction of vertices $v \in V$ with degree $d(v) = d$. For different d_1, d_2, \dots, d_n , the collection $\{f(d_1), f(d_2), \dots, f(d_n)\}$ is called the degree distribution of G . A useful generalization of degree is the notion of vertex strength, which is obtained simply by summing up the weights of edges incident to a given vertex. The distribution of strength is sometimes called the weighted degree distributions defined in analogy to the ordinary degree distribution. In this work

we will use both vertex degree and vertex strength as network attributes for classification. Specifically for each vertex v (for example representing a subject labeled as “Case”) the following attribute will be obtained.

1. The degree of v with respect of the set of instances with in the same class (label) of v (i.e. the number of *cases* connected to v)
2. The strength of v with respect of the set of instances within the same class (label) of v (i.e. the number of case instances/subject connected to v)
3. The degree of v with respect of connected instances with different label of v (i.e. the number of *control* instances/subject connected to v)
4. The strength of v with respect of connected instances with different label of v (i.e. the number of *control* instances/subject connected to v)

2.5 Dissimilarity-based attributes

In order to extend the network statistic reported above we introduce a set of dissimilarity attributes. Dissimilarities have been used in *pattern recognition* for many years, leading to many different known algorithms and important questions. For example, the idea of “template matching” is based on dissimilarities: objects are given the same class label if their difference is sufficiently small (Duda, Hart, & Stork 1973). This is identical to the nearest neighbor rule used in vector spaces (Duda, Hart, & Stork 2001). Also many procedures for cluster analysis make use of dissimilarities instead of the standard features space representation (Theodoridis & Koutroumbas 2006). A detailed description, providing mathematical foundation, designed procedures, and real world examples for building pattern recognition systems based on dissimilarity representation may be found in (Pekalska & Duin 2005).

As observed above, a typical way of representing instances is through the use of a vector of available chosen attribute values (e.g. clinical data in BD). Here, we extend the set of standard and network attributes by using a *dissimilarity representation* which can express, through a function $D(x, y)$, the dissimilarity between e.g. the clinical measurements for the pair of patients x and y . By extending $D(x, y)$ for all pairs (of patients), we can formulate a dissimilarity matrix whose rows can be assessed also by representing any patient $x \in X$ through the mapping $(X, P) \rightarrow R$, defined as $\varphi(x, P) = [D(x, y_1), D(x, y_2), \dots, D(x, y_n)]$, where X and P respectively denote a set of *case/control patients* and a set of n *prototype patients*. Here the difference between X and P reflects the need to discriminate case/control patients in X as compared to a common set of n prototype patients in P . In particular we considered the following attributes.

1. For each pair of patient the difference between their (standard) clinical covariates.
2. For each pair of patient the difference between their network statistics (i.e. the features considered in the above section)

3 NUMERICAL RESULTS

In this section we report the accuracy of the CC paradigm when applied to the set of data described previously. In particular, the goal of our investigation is to check whether by adding network and differential information, we can improve the accuracy of a classification problem. Numerical experiments are evaluated using standard indexes such as sensitivity, specificity, Positive Predictive Value (PPV), Negative Predictive Values (NPV) and accuracy. These indexes measure the capability of an inference system to classify entities in e.g. “case vs control” classification problem (Florkowski 2008). To avoid biased estimation we sub-sampled the available data (10-fold cross validation). As reported above the performances are given for a base model (a standard inference method, i.e. PLS) common for all the input attributes in such a way that we can set up homogeneously our numerical experiments. As it is shown in Table 1, all features passed to the classification models provide very good accuracies $\geq 90\%$. Importantly, features using the network and dissimilarity information outperformed in general the results obtained using only standard-based information, i.e. clinical covariate (please note the two marked values). Moreover, in order to synthesize the results, we grouped the features as follows.

1. STA. Only standard attribute are considered.
2. STA DYN FEA. Standard attribute and network statistics.
3. STA DIF-OVER-STA FEA. Standard and dissimilarity based attributes. Dissimilarities are computed using standard attribute values.
4. STA DIF-OVER-DYN FEA. Standard and dissimilarity based attributes. Dissimilarities are computed using network attribute values.
5. STA DYN DIF-OVER-STA FEA. Standard, network and dissimilarity based attributes. Dissimilarities are computed using standard attribute values.
6. STA DYN DIF-OVER-DYN FEA. Standard, network and dissimilarity based attributes. Dissimilarities are computed using network attribute values.
7. STA DYN DIF-OVER-STA DIF-OVER-DYN FEA. Standard, network and dissimilarity based attributes. Dissimilarities are computed both over network and dissimilarity attribute values.

Feature types	Sen	Spe	PPV	NPV	Acc	class
STA	0.95	0.9	0.904762	0.947368	0.925	S
STA_DYN_FEA	1	0.95	0.952381	1	0.975	SD
STA_DIF-OVER-STA_FEA	0.975	0.925	0.928571	0.973684	0.95	SF
STA_DIF-OVER-DYN_FEA	0.975	0.95	0.95122	0.974359	0.9625	SF
STA_DYN_DIF-OVER-STA_FEA	0.975	1	1	0.97561	0.9875	SDF
STA_DYN_DIF-OVER-DYN_FEA	0.9	1	1	0.90909	0.95	SDF
STA_DYN_DIF-OVER-STA_DIF-OVER-DYN_FEA	1	0.975	0.97561	1	0.9875	SDF

Figure 1. Classification Performances.

class	Sen	Spe	NPV	PPV	Acc
S	0.95	0.9	0.947368	0.904762	0.925
SD	1	0.95	1	0.952381	0.975
SDF	0.958333	0.991667	0.961567	0.99187	0.975
SF	0.975	0.9375	0.974022	0.939895	0.95625

Figure 2. Classification Performances by class of attribute Types.

With the above clustering, we averaged the performances obtained by using the features considered in each cluster expressed above, in order to summaries the results when we add to the standard information (standard features) either network or differential information. As reported in Fig. 2 we can observe that the input which contain both networks and differential information when added to the standard clinical attributes increase on average the performances. This way providing a general better accuracy.

4 CONCLUSIONS

Network technologies and theoretical methods for mining complex structures are now fundamentals at each conceptual level. They must proceed together if new properties and interesting knowledge from complex networks are to be observed. Our task here was to provide the reader with an overview of this research area by introducing the way on how information is managed (avoiding to assume i.i.d. samples) to mining new knowledge for supporting diagnostic decision. To this aim, following an experimental investigation we showed the reader how different kind of information can provide a benefit for a classification problem. The main issue of this investigation was to show a case where the accuracy of the classification may increase by adding both network and dissimilarity information. Clearly, in order to give a significant evidence of the usefulness of this approach, more data and models have to be compared through suitable statistical tests, aiming to take into account the not so straightforward applicability of the required statistical assumptions for the machine learning algorithms, see for instance the book (Japkowicz & Shah 2011). This is a first extension to this work

which we are immediately interested in our future analysis. Moreover, referring to the information introduced through the dissimilarity features, we must keep in mind that the choice of a correct prototype set can be critical and has to be carefully considered. This will be another question which we will be considered in our future study.

The authors declare that there is no conflict of interests regarding the publication of this paper.

REFERENCES

- Bansal, N., A. Blum, & S. Chawla (2002). Correlation clustering. *Machine Learning*, 238–247.
- Bollobás, B. (1998). *Modern Graph Theory*. Springer-Verlag.
- Castelnuovo, G. (2010). No medicine without psychology: the key role of psychological contribution in clinical settings. *Front Psychol* 1(4), 3–11.
- Castelnuovo, G., G. Manzoni, G. Pietrabissa, S. Corti, E. Giusti, E. Molinari, & S. Simpson (2014). Obesity and outpatient rehabilitation using mobile technologies: the potential mhealth approach. *Front. Psychol.* 5 (559).
- Castelnuovo, G. & S. Simpson (2011). Obesity-e-health for obesity-new technologies for the treatment of obesity in clinical psychology and medicine. *Front. Psychol.* 7, 5–8.
- Cava, C., I. Zoppis, M. Gariboldi, I. Castiglioni, G. Mauri, & M. Antoniotti (2013). Copy-number alterations for tumor progression inference. In N. Peek, R. Marn Morales, and M. Peleg (Eds.), *Artificial Intelligence in Medicine*, Volume 7885 of *Lecture Notes in Computer Science*, pp. 104–109. Springer Berlin Heidelberg.
- Diestel, R. (1997). *Graph Theory*. Number 173 in Graduate Texts in Mathematics. Springer.
- Duda, R., P. Hart, & D. Stork (1973). *Pattern classification and scene analysis*. Wiley New York.
- Duda, R., P. Hart, & D. Stork (2001). *Pattern Classification*. Wiley.
- Florkowski, C. (2008). Sensitivity, specificity, receiver-operating characteristic (roc) curves and likelihood ratios: Communicating the performance of diagnostic tests. *The Clinical Bio-chemist Reviews* 29, S83–S87.
- Getoor, L. & B. Taskar (2007). *Introduction to Statistical Relational Learning (Adaptive Computation and Machine Learning)*. The MIT Press.
- Gross, J. & J. Yellen (2005). *Graph Theory and Its Applications, Second Edition (Discrete Mathematics and Its Applications)*. Chapman & Hall/CRC.
- Indyk, W., T. Kajdanowicz, P. Kazienko, & S. Plamowski (2012). Map reduce approach to collective classification for networks. In *Proceedings of the 11th International Conference on Artificial Intelligence and Soft Computing—Volume Part I, ICAISC’12*, Berlin, Heidelberg, pp. 656–663. Springer-Verlag.
- Japkowicz, N. & M. Shah (2011). *Evaluating Learning Algorithms: A Classification Perspective*. New York, NY, USA: Cambridge University Press.
- Jensen, D., J. Neville, & B. Gallagher (2004). Why collective inference improves relational classification. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’04*, New York, NY, USA, pp. 593–598. ACM.
- Kolaczyk, E. (2014). *Statistical analysis of network data with R*. Springer Link: Bücher. Springer.
- Lamprinos, I., C.P.C., H. Schmuhl, H. Demski, C. Hildebrand, & M. Plossnig (2014). Mobile personal health application for empowering diabetic patients. *Journal of the International Society for Telemedicine and eHealth* (21), 3–11.
- Macskassy, S. & F. Provost (2007, December). Classification in networked data: A toolkit and a univariate case study. *J. Mach. Learn. Res.* 8, 935–983.
- Mangasarian, O.L. & W.H. Wolberg (1990). Cancer diagnosis via linear programming. *23(5)*, 1–18.
- Marsland, S. (2011). *Machine Learning: An Algorithmic Perspective*. CRC Press.
- Mitchell, T. (1997). *Machine Learning*. New York, NY, USA: McGraw-Hill, Inc.
- Newman, D., S. Hettich, C. Blake, & C. Merz (1998). Uci repository of machine learning databases.
- Pekalska, E. & R. Duin (2005). *The Dissimilarity Representation for Pattern Recognition: Foundations And Applications (Machine Perception and Artificial Intelligence)*. River Edge, NJ, USA: World Scientific Publishing Co., Inc.
- Santoro, E. (2013). Social media and medical apps: how they can change health communication, education and care. *Recenti Progressi in Medicina* 104(5), 179–180.
- Santoro, E., P. Caldarella, & A. Villella (2011). Using web 2.0 technologies and social media for the cardiologist’s education and update. In *G Ital Cardiol (Rome)*, Volume 12, pp. 174–181.
- Santoro, E. & G. Quintaliani (2013). Using web 2.0 technologies and social media for the cardiologist’s education and update. In *G Ital Nefrol*, Volume 30.
- Theodoridis, S. & K. Koutroumbas (2006). *Pattern Recognition, Third Edition*. Orlando, FL, USA: Academic Press, Inc.
- Witten, I., E. Frank, & M. Hall (2011). *Data Mining: Practical Machine Learning Tools and Techniques: Practical Machine Learning Tools and Techniques*. The Morgan Kaufmann Series in Data Management Systems. Elsevier Science.
- Zoppis, I. & G. Mauri (2008). Clustering dependencies with support vectors. *Lecture Notes in Electrical Engineering* 6, 155–165.
- Zoppis, I., D. Merico, M. Antoniotti, B. Mishra, & G. Mauri (2007). Discovering relations among go-annotated clusters by graph kernel methods. Volume 4463 of *Lecture Notes in Computer Science*, pp. 158–169. Springer.

VoIP providers trunk round trip delay remote measurement methodology

Martin Mikulec, Jan Rozhon & Miroslav Voznak

*Department of Telecommunications, Faculty of Electrical Engineering and Computer Science,
Technical University of Ostrava, Ostrava, Czech Republic*

ABSTRACT: The paper aims at round trip time measurement methodology. No special hardware or software needs to be implemented into Operators infrastructure. The measurement is based on packet voice payload comparison and round trip delay evaluation from timestamps. The methodology can be applied as an external monitoring tool of VoIP Operator infrastructure trunk lines or call admission control functions.

Keywords: voice over ip, round trip delay, speech quality, measurement, methodology

1 INTRODUCTION

Nowadays, the Voice over Internet Protocol (VoIP) is widely used both in the large networks as well as in SOHO (Small Office, Home Office) networks. This reason is that the audio or video packet data communications bring a consolidation of transmission networks which means economic benefit for network providers and end users. The most IP Telephony Operators reduce the VoIP payload the way that the stream is separated from the rest of the network using virtual LANs (Local Area Networks), even though, it is necessary to monitor the quality of calls on each node. Measuring the quality of lines is important but often overlooked part not only of IP telephony. At the beginning of the VoIP service development, there were many quality of service issues (Pravda & Vodrazka 2007), which were step by step resolved by additional recommendations. However not all recommendations are implemented by VoIP Operators and end users, so there is still necessary to monitor VoIP infrastructure by external monitoring tools. The operator can monitor VoIP infrastructure by hardware or software tools implemented inside its infrastructure. There are several tools available for this purpose. When the Operator wants to monitor its infrastructure remotely as a service, the monitoring service provider needs to implement monitoring probes, which send results to the monitoring center or dedicate part of its infrastructure for service calls, which will be used to provide necessary data to the monitoring center. This paper offers a solution, how to monitor VoIP infrastructure remotely without any hardware or software implementation into VoIP Operators infrastructure.

2 STATE OF THE ART

Actual methods of measuring a quality of the lines between the VoIP Operators are based on the special software or hardware solutions implemented inside the operators infrastructure. Our laboratory has designed installable solution based on establishing VoIP calls between installed probes (Rezac, Rozhon, Slachta, & Voznak 2015). The measurement is originated in regular time intervals, and the line quality is evaluated after each call. The Perceptual Evaluation of Speech Quality methodology is used to evaluate the quality of the calls. The results are represented in MOS (Mean Opinion Score) scale as defined by the ITU-T recommendation P.800 (ITU-T 1996). The results are presented in the form of a map where peaks represent the probes, and the measurement paths are shown in the form of evaluated edges. There are lots of reasons why not to install mentioned probes into operators VoIP infrastructure. The operator could find inappropriate to install external software solution into its infrastructure because of the security policies. Moreover, when the monitoring should be provided as an external service to the Operator, there is necessary to open particular connections through the firewall to ensure getting required files for speech quality evaluation. Last disadvantage of the method is non real-time evaluation of the quality after finishing the testing call.

These three aspects were motivation for designing a new methodology for line quality monitoring based on the real-time evaluation by the external tool using existing VoIP Operator infrastructure. The new method is using outgoing and incoming packet evaluation. The packets are sent from

evaluation system, transmitted through VoIP Operator infrastructure via common extensions and sent back to the evaluation system. According to the fact, that every incoming packet is evaluated, the system can provide results in real-time mode. The measuring system is described in following chapters in detail (Chromy, Jadron, Kavacky, & Klucik 2013). Real-time speech quality assessment is important in many types of networks and one of such examples is the applied evaluation in Mobile ad-hoc networks in order to maximize the number of acceptable VoIP calls (De Rango, Fazio, Scarcello, & Conte 2014). The proposed methodology in this paper can be applied for video quality assessment as well due to capability of delay and packet loss measurement which affect overall quality of experience (Uhrina, Hlubik, & Vaculik).

3 MEASUREMENT PRINCIPLES

The new methodology evaluates outgoing and incoming RTP (Real Time Protocol) packets. The packets bring test VoIP call, which is established between Evaluation System and measured VoIP Operators networks. The Fig. 1 depicts the packet flow from Evaluation System over measured networks and back to the Evaluation System.

The Evaluation System generates test call, which is sent over SIP trunk to the VoIP Operator #1 to the specific pre-arranged transfer extension number. The VoIP Operator #1 transfers the call through established SIP trunk to the specific VoIP Operator #2 extension number with pre-arranged echo function. Operator #2 returns the incoming RTP packets back to the VoIP Operator #1 and the packets go back to the Evaluation System with appropriate delay. The measurement methodology prerequisite arranged SIP trunk connection between the systems and preconfigured extension

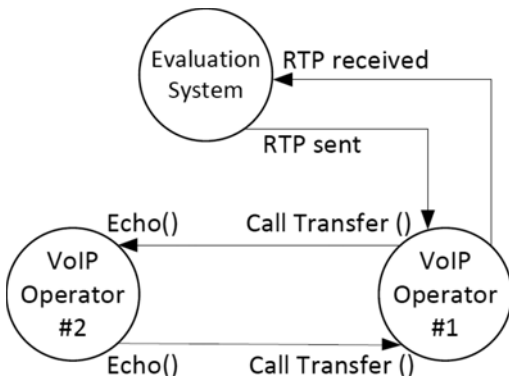


Figure 1. RTP packet measuring scheme.

numbers with transfer and echo functions. The transfer and echo functions are essential functions, which every modern switching system should offer to its customers. The testing call consists of definite numbers of RTP packets. The number of packets is related to the payload size of the codec. The G.711 a-law codec used in our measurement has 20 ms payload size, therefore, the number of packets will be 50 packets for every second of the call. Every outgoing and incoming RTP packet is captured at the particular interface and send to the measurement algorithm with following steps:

1. Read the RTP packet,
2. Save timestamp of the RTP packet,
3. Encode the packet into hex-binary form,
4. Cut first 24 characters to obtain only RTP payload,
5. Save RTP payload and timestamp to packet dictionary.

Every next outgoing or incoming packet is captured, encoded and compared with the already stored packet in the dictionary. When the dictionary does not contain the packet with the same payload, the packet is added to the dictionary with the particular timestamp. When the dictionary contains the packet with the same payload, it means this the packet is the answer to the send packet. The difference between the packets timestamps is calculated and stored into result dictionary with the timestamp of the received packet. Then the payload is removed from the payload dictionary as is depicted in Fig. 2.

The resulting dictionary contains round-trip delay between Evaluation System and VoIP Operator #2 for every pair of packets with the same payload. From the practical reason, there is useful to evaluate average round-trip delay time after a specific period or after the testing call according to the requirements. There are all necessary data

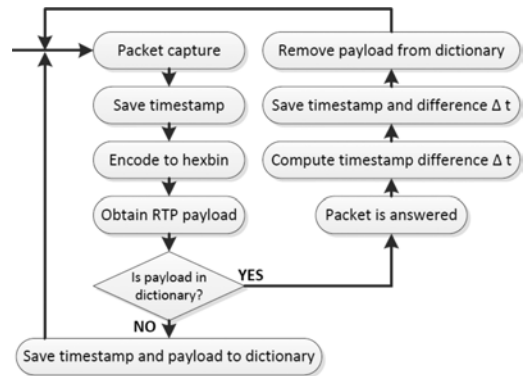


Figure 2. Measuring algorithm.

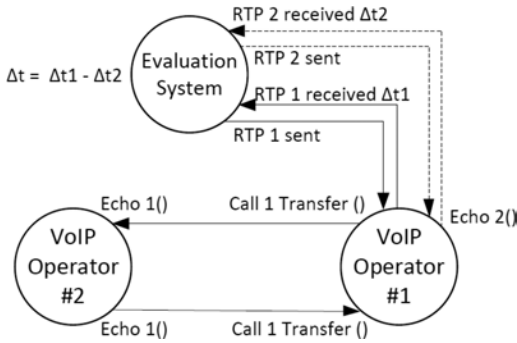


Figure 3. RTP packet measuring scheme with Echo.

in the result dictionary both for real-time or after call evaluation. The packet dictionary contains a list of unanswered packets. When the packet dictionary is empty, we can consider the line without any packet loss. Unanswered packets mean either the packets are waiting for an answer packet, or the answer packet was lost. According to the fact, that the number of packets per one second is small (50 pps), every lost packet generate huge packet loss. Therefore, the real-time evaluation is inaccurate, the evaluation is meaningful for 10 and more second long calls. In ideal network conditions, we can consider the round-trip delay between Evaluation System and VoIP Operator #2 as the same as the round-trip delay between the VoIP Operator #1 and VoIP Operator #2. In a real scenario, we have to calculate with the round-trip delay between Evaluation System and VoIP Operator #1 (Δt_2) and subtract the delay from the overall round-trip delay between Evaluation System and VoIP Operator #2 (Δt_1). The final round-trip delay (Δt) can be expressed as $\Delta t = \Delta t_1 - \Delta t_2$. The real scenario with particular sent and received RTP streams between operators is depicted in Fig. 3.

4 MEASUREMENT REALIZATION

The measurement system is based on the open source Asterisk PBX soft switch, which generates test call and ensures connectivity to another VoIP Operators. The sent RTP packets are sniffed on the particular interface by Scapy (Biondi 2010). Scapy is a Python program which enables the user to send, sniff and dissect and forge network packets. Every packet is evaluated inside a Python script, stored into the particular dictionary and finally the results are displayed after a set time, see Fig. 4. We have tested the measurement in test scenario in our laboratory between three Asterisk PBX servers in the following measuring scheme as is depicted in Fig. 5.

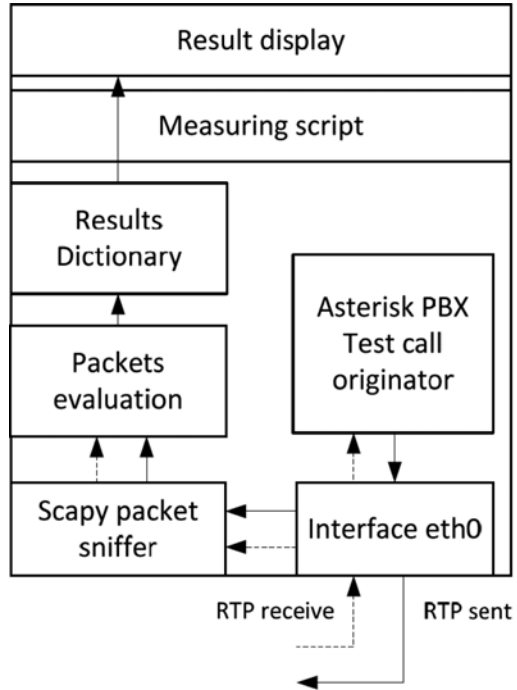


Figure 4. Evaluation system scheme.

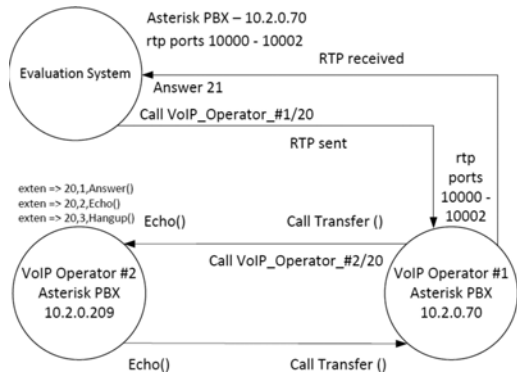


Figure 5. Test scenario scheme.

4.1 Evaluation system asterisk PBX

Asterisk PBX (Bryant 2013) provides test call generator and connection with another VoIP Operators. The system strictly requires G711 a-law codec setting between measured operators. The G.711 a-law codec can be set in the sip.conf configuration file in particular trunk setting part via subsequently placed parameters *disable = all* and *enable = alaw*.

The sip.conf configuration file is used to create SIP trunk between Asterisk in Evaluation System

and VoIP Operator. The setting is standardized according to RFC 3261 and it needs to be negotiated between Evaluation System and VoIP Operator (Rosenberg, Schulzrinne, Camatillo, Johnston, & Peterson 2002) and (Chamraz & Baronak 2015). Another setting is realized in Asterisks dialplan in extension.conf configuration file. The call is established via extension 20 and when VoIP Operator accepts the call, the extension 21 with playback record is called.

```

exten = 20, 1, Answer()
exten = 20, 2, Dial(SIP/VoIP_Operator/20, L(20000))
exten = 20, 4, Hangup()

exten = > 21, 2, Answer()
exten = > 21, 3, Playback(sound_sample)
exten = > 21, 5, Hangup()

```

The call can be originated remotely from Python script by calling Asterisk CLI command.

```

p = subprocess.Popen(["/usr/sbin/asterisk", "-rx",
"channel originate Local/21@default extension 20"],
stdout = subprocess.PIPE).

```

4.2 Scapy packet sniffer

When the test call is established, the Scapy packet sniffer will sniff every packet on the defined network interface which corresponds to a predefined filter. The filter needs to be set according to the real IP addresses and used port to ensure, that only test call packets will be filtered and sent to the pkt_callback script to packets evaluation. We can also limit the number of sniffed packets or run time in seconds. In our test scenario, we have used following scapy settings, related to used RTP ports.

```

sp.sniff(iface = "eth 0", prn = pkt_callback, filter =
"udp and ((src port 10000 and dst port 10002) or
(src port 10002 and dst port 10000) or
(src port 10000 and dst port 10000) or
(src port 10002 and dst port 10002)) and
((dst host 10.2.0.51 and src host 10.2.0.70) or
(dst host 10.2.0.70 and src host 10.2.0.51))",
store=1, count =10000, timeout = 60).

```

4.3 Packet evaluation

Every sniffed packet is sent to evaluation script which runs evaluation processes according to the

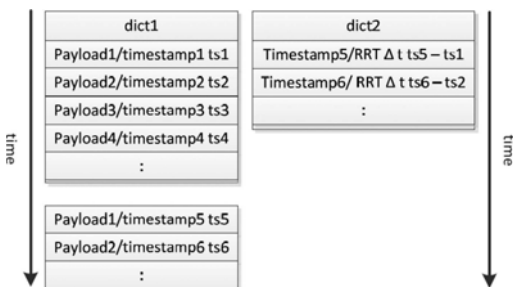


Figure 6 Dictionaries.

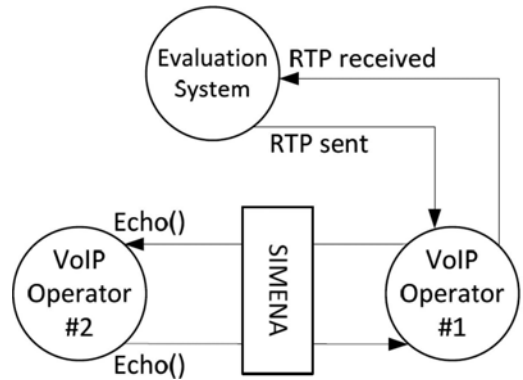


Figure 7. Measurement verification scheme.

scheme in Fig. 2. The algorithm is presented in the following code.

```

def pkt_callback(pkt):
a = pkt.getlayer(sp.Raw).load
d = a.encode('hex')
b = pkt.time/*packet timestamp t1 storing
a = d[24:]/* getting RTP payload
if dict1.has_key(a):/* if in dictionary
c = dict1[a]/* get t1 timestamp
deltat = b-c/* get round trip delay.
dict2[c] = deltat;/* save result to dict.
del dict1[a]/*remove payload from dict.
else:/* if payload is not in dictionary.
dict1[a] = b/* store payload and timestamp.

```

Figure 6 present the packets sent to dictionary 1 and dictionary 2 according to the code above. The result round trip delay can be obtained from dictionary 2 according to measuring needs. In our script, we have preferred to obtain average round trip value every one second.

5 MEASUREMENT VERIFICATION

We have tested our measurement system through SIMENA network emulator, which emulated latency and packet loss between VoIP Operators. The SIMENA emulated latency and packet loss between VoIP Operator #1 and VoIP Operator #2 according to the following scheme in Fig. 7. The control measurements confirmed theoretical assumption. The set parameters on the line between VoIP Operators were verified by Evaluation System.

6 CONCLUSION

We have verified by laboratory measurement that our new methodology based on packet evaluation of RTP payload is suitable for line quality monitoring between VoIP Operators. The Evaluation System based on open source tools can send service call

to VoIP Operator infrastructure and evaluate packet stream, which returns from measured infrastructure. The round trip delay and packet loss can be evaluated by Evaluation System. Contribution of this research lies in a new proposal how to measure especially round-trip delay, nevertheless the proposed methodology enables packet loss measurement as well, without necessity of internal implementation of measurement apparatus in network of cooperating operators. The proposed approach is based on external measurement and using solely *echo* function in systems where the call is terminated. The paper provides detail explanation of principles and implementation of the proposed methodology.

ACKNOWLEDGMENT

This research was supported by the SGS grant No. SP2016/170, VSB-Technical University of Ostrava, Czech Republic.

REFERENCES

- Biondi, F. (2010). Scapy v2.1.1-dev documentation. Technical report, Scapy Community.
- Bryant, R. (2013). *Asterisk: the definitive guide. Fourth edition*. Sebastopol: O'Reilly.
- Chamraz, F. & I. Baronak (2015). Impact of admission control methods to the traffic management. *Advances in Electrical and Electronic Engineering* 13(4), 280–288.
- Chromy, E., M. Jadron, M. Kavacky, & S. Klucik (2013). Admission control in ims networks. *Advances in Electrical and Electronic Engineering* 11(5), 373–379.
- De Rango, F., P. Fazio, F. Scarcello, & F. Conte (2014). A new distributed application and network layer protocol for voip in mobile ad hoc networks. *IEEE Transactions on Mobile Computing* 13(10), 2185–219.
- ITU-T (1996). Recommendation p.800: Methods for objective and subjective assessment of quality. Technical report, Geneva.
- Pravda, I. & J. Vodrazka (2007). Voice quality planning for ngn including mobile networks. *IFIP International Federation for Information Processing* 245, 376–383.
- Rezac, F., J. Rozhon, J. Slachta, & M. Voznak (2015). Speech quality measurement in ip telephony networks by using the modular probes. *Communications in Computer and Information Science* 522, 172–181.
- Rosenberg, J., H. Schulzrinne, G. Camatillo, A. Johnston, & J. Peterson (2002). Sip: Session initiation protocol. Technical report, IETF: Network Working Group, Internet Society, USA.
- Uhrina, M., J. Hlubik, & M. Vaculik (2013). Correlation between objective and subjective methods used for video quality evaluation. *Advances in Electrical and Electronic Engineering* 11(2), 135–146.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Automatic identification and data capture techniques by radio frequency identification RFID tags applied to reader authentication

H. Saadi & R. Touhami

LINS, Faculty of Electronics and Informatics, USTHB, Bab Ezzouar, Algiers, Algeria

M.C.E. Yagoub

School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Ontario, Canada

ABSTRACT: Radio Frequency Identification (RFID) is an emerging technology, with applications already mastered and spread: animal identification, access control, control of logistics flows, information from environmental sensors, etc. However, lack of security and some environmental issues can significantly deteriorate the performance of such systems. Therefore, this paper investigates different areas of application in which RFID technology, as data capture technique, presents some drawbacks that must be resolved by authentication techniques and algorithms, and the risk of applying such techniques according to international standardization.

1 INTRODUCTION

Radio-Frequency Identification, or RFID, is a method to memorize and recover remote data by using markers called “radio-labels” (“RFID tag” or “RFID transponder”).

A typical RFID system is composed of a reader, tags and data management system, see Figure 1.

Tags are small objects, such as auto adhesives labels, which can be stuck or incorporated into objects or products or even be established in living organisms (animals/human bodies).

They include an antenna associated with a microchip, which enables them to receive and answer the radio requests emitted by the reader (Finkenzeller) 2003.

This technology can be used to identify different targets like:

- persons while being integrated in passports or payment cards (as contactless cards);
- pets (cats, dogs) whose identification is obligatory in many countries, while being established under the skin (implemented as subcutaneous chip).



Figure 1. Typical RFID system.

2 PRINCIPLE

A radio-identification system is activated by an electromagnetic transfer of energy between a reader and one or more tags. A tag is composed of a chip and an antenna (see Fig. 2) (Finkenzeller) 2003.

2.1 Readers

Readers are active devices transmitting radio frequencies to activate neighbor tags.

RFID systems generate and reflect electromagnetic waves; therefore, as radio systems, they are subject to strict regulation. RFID systems have to ensure not to disrupt the operation of other radio systems working in or close to their operating frequency range. Such as TV, emergency services, maritime and air radio services, mobile phones, etc.

Table 1 summarizes the different frequencies used in RFID systems.

The frequency used is variable, according to the type of application concerned and the required performances (Finkenzeller) 2003. We can have:

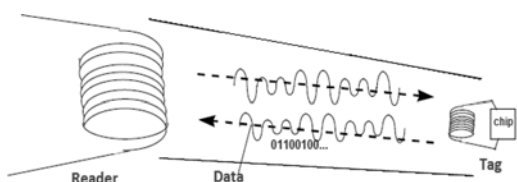


Figure 2. Functioning principle of RFID system (Finkenzeller) 2003.

Table 1. Main frequencies used in RFID systems.

Spectrum frequency	RFID frequency	Type of tag
LF	125–134.2 kHz	Passive
HF	13.56 MHz	Passive
VHF	868 MHz (Europe) 915 MHz (USA)	Passive/active
UHF	2.45 GHz	Active
SHF	5.8 GHz	Active

- 125 kHz for low frequency applications;
- 134.2 kHz for transponder load in the case of an FSK transmission (Texas Inst. Series) 2000;
- 13.56 MHz (ISO 14 443 A 1-4, ISO 14443B 1-4, ISO 15693-3 and ISO 18000-3), currently most widespread in industry and general public for applications with limited range (badges, building access);
- 865 MHz to 868 MHz in the European Union (EPCglobal and ISO 18000-6c; the frequencies and transmitting powers depend on legislations);
- 915 MHz in the United States;
- 2.45 GHz or 5.8 GHz, used, for example, for wireless payment.

High frequencies have the advantage of allowing an exchange of information (between reader and tag) with high bit rate of data, more significantly than at low frequencies. This significant bit rate allows the implementation of new functionalities within the tags (cryptography, large memory, anti-collision). On the other hand, a lower frequency will profit from a better penetration in the material.

The anti-collision approach is the possibility for a reader of dialoguing with a tag when more than one tag is present in his area of detection or in its area of reading. Several anti-collision algorithms are described by standards (ISO 14443, ISO 15693 and ISO 18000), and several research has been done in this aim.

2.2 Tags

They are passive devices, not requiring any source of energy apart from that provided by the readers at the time of their interrogation. Previously, the reading of the passive chips was limited to a distance from about 10 meters, but nowadays, thanks to the technology used in the communication systems with remote space, this distance can be extended to about 200 meters.

In addition to energizing the tag, the reader sends a particular signal of interrogation, for which the tag answers.

One of the simplest ways to answer is the reference of a numerical identification, for example that of the standard EPC-96, which uses 96 bits.

A table or database can be consulted to ensure an access, counting or data tracking on an assembly line.

The tag could be extremely discrete by his smoothness, his reduced size (few mm²), and its negligible mass. Its cost has become effective. The tag is mainly composed of

- an antenna;
- a chip;
- a substrate and/or an encapsulation.

Let us note also the use of active and semi-active tags (also called BAP: Battery-Assisted Passive tags) and passive tags assisted by battery which include a battery.

Active tags are equipped with a battery enabling them to emit a signal. So they can be read since long distances, contrary to passive tags. However, an active data transmission highlight to all the surrounding environment about its presence, thus raising questions about safety.

Semi-active tags do not use their battery to emit signals. They act like passive tags on the communication level. However, their battery allows them, for example, to record data during transport. These tags are used in the sending of products under controlled temperature and record the temperature of the goods at regular intervals.

We have also tags without chips to be used as single identifier. With a very weak cost level, these

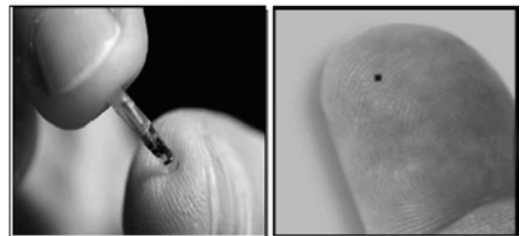


Figure 3. Tag designed to be injected under skin of animals (Perret) 2014.



Figure 4. RFID tags with different shapes (Perret) 2014.

last can be alternatives to code-bars and can be implanted on human body (Perret) 2014 as illustrated in Figure 3.

According to that and based on the targeted application, we can find different shapes of tags as shown in Figure 4.

3 ETHICS, PRIVATE LIFE AND AUTHENTICATION

3.1 *In the world*

In the years 2000, in all developed countries, RFID chips have been quickly standardized. In 2010, the establishment of micro cells in the human body was initiated (VeriChip chip or “human code bars”), with the correlative risk of the shapes of control of individuals (Perret) 2014.

This was done before even having a clear legislation and in-depth ethical debates, in particular concerning the active or passive devices and increasingly miniaturized (in 2006, Hitachi proposed a square chip of $0.15 \times 0.15 \text{ mm}^2$, smaller than the diameter of some hairs (Hitachi) 2006.

Inserted under the skin (Perret) 2014, or put on cloths (wearable computing or cyber-clothing), RFID communicating objects are becoming part of our daily life; a German company, Ident Technology, developed devices working under human or animal skins (GEESNT) 2005. These chips are, as many other similar innovations, sources of ethical questions and risks of new drifts (COE) 1997, (Nsanze) 2005.

3.2 *In Europe*

After a 2005 report about new implants in the human body (EC) 2009 and a round table organized by the GEE (European Group of Ethics of Sciences and the New Technologies) (JO) 2002, the European commission asked the BEPA (Office of the Advisers of European Policy) to get their opinion. On March 2015, on the request of the GEE, an ethical report called “Aspects of the implants TIC in the human body” was published (GEESNT) 2005.

The principal concerned rights are mainly about human dignity, right to the integrity of the person, and data protection of personal information (Recette) 1990.

The question also touches the public health, the protection of the private life in the electronic sector of communications (COE) 1997, the legislation on medical implantable active devices (UNESCO) 1997, the approval and the right to access information (Convention) 1997, the protection of people against automated data processing of personal

data matter (Point) 2003, as well as any possible abusive uses that can result (European Recommendation) 2013.

In May 2009, the European Commission published a recommendation (European Recommendation) 2013, centered on the systematic deactivation of RFID tags at retailing points. For applications that did not decontaminate systematically the tags, the start of an RFID application is subjected to the evaluation of impacts on private life (EIVP or Privacy Impact Assessment PIA).

In July 2014, a European standard was published (INTO 16571) giving the methodology to be followed to carry out an EIVP. The EIVP report must be transmitted to the organization in charge of data protection involving personal information (GEESNT) 2005.

4 OBSTACLES

4.1 *Metallic environment*

The reading of radio-tags posed on objects located in or close to metallic containers is more challenging. Because of the presence of a ground plan, the performance of the antenna of the tag is modified. This can reduce in a drastic way the reading distance.

New families of tags have been then recently developed, focusing on the design of the antenna, in order to integrate the presence of a metal plan, making it possible to keep long reading distances. In all cases, a tag placed inside a metal enclosure could not be read by a reader located outside. It is the effect of “Faraday caging”, which carries out an electromagnetic shielding.

4.2 *Collisions*

When several tags are in the area of reading of the same reader, the communication is scrambled by the simultaneous activity of the tags.

The collision detection is in fact a detection of transmission error, using several methods. As soon as an error is detected, the anti-collision algorithm is applied.

Several anti-collision methods were developed. They can be divided into random methods and deterministic ones, and according to the kind of multiple access to the communication channel, they can be classified again into four principal methods (Myung et al., 2006), (Vogt, 2002), (Lee et al., 2005), (Hu, 2008), and (Saadi, 2011):

- The frequency method FDMA: Each tag communicates on different frequency ranges with the reader. In practice, it is unusable on a large scale.

- The space method SDMA: With a directional antenna and variable powers, the reader can gradually cover each part of its space to communicate with each tag and to inhibit it, while waiting to reactivate it for initiating the data transfer. In practice, the presence of tags separated by a small distance makes this method ineffective.
- The temporal method TDMA: The reader proposes to the tags a series of channels of time, also called slots, in which they can answer. The tags randomly choose the slot of time in which they will answer. If a tag is the only one to answer in this slot of time, it is detected and inhibited by the reader. If there are several tags which answer at the same time, it will be necessary to carry out this method again. Gradually, all the tags are known and inhibited; it is then enough for the reader to reactivate the tag with whom it wishes to communicate. In practice, the random side makes that the duration of this technique uncontrollable.
- The systematic method: There are many researches describing the systematic methods. This method consists in detecting and then inhibiting all the tags, in turn, by completing the path of all the possibilities of identifiers (for example, the reader sends a request like “all the tags whose first bit of identification is 1 must respond”; thus, if only one tag appears, the reader inhibits it, and then switches to the tags with first bit 0, and so on). In practice, this method can be sometimes quite long to be completed.
- Access control to sensitive buildings in which the system of radio-identification replaces the magnetic badges, allowing an authentication of people without contact.

Most RFID badges allow only the use at very limited distance (few centimeters), but they have the advantage of allowing a read-write of the chip to memorize information (biometric, for example), see Figure 5.

- Distant traceability of objects (fixed or mobiles): for example, palettes or containers can be followed in warehouses through UHF tags.
- Reading through (and thus the human body). In 2008, the RFID journal Awards went to the Omni-ID company that presented an RFID tag readable through water and near metal, with a high level of confidence of 99.9%.
- Microwave tags working at 2.45 GHz allow long distance access control of vehicles, as in large industrial parks. These tags are generally active tags.
- Traceability of food: In the chain of the cold, food can theoretically be followed by a chip recording the variations in temperature.

5 USE OF RFID

5.1 Marking of objects

This can involve:

- Implementation of a system of identification and memorization: in a current way, low frequency devices (125 to 135 kHz) are used for the traceability of objects. The traceability of objects such as books in a library or localization of a luggage in an airport;
- A less known usage of the RFID technology, but tends to grow, is related to the rational management of domestic waste, in order to set up an inciting tariffing (le RFID) 2014.
- Access control: done by vicinity badges or “free-hands” that allow a use until approximately 150 cm. They can contain a numerical identity or an electronic certificate, which permit an access to an object or its activation. Used, for example, for the access control of public transportation systems.
- Electronic keys allowing protection “without lock” of buildings or vehicle doors.

5.2 Financial transactions

This can involve the systems of contactless payment with credit cards, mobile phones ... It uses the near field communication properties to make secure payments. An integrated chip and an antenna make it possible to consumers to pay with their contactless cards on a reader at a point of sale. At Hong Kong and in Holland, tags in the shape of credit card are widespread like electronic means of payment. They are also used as transport titles on public transportation networks.

5.3 Marking of human, animals and plants

This involves:

- Identification of plants, wild animals or pets like cats and dogs. The tag is usually installed under



Figure 5. Access control with RFID.

the skin in the neck. It is generally done with a low frequency tag (125 to 135 kHz).

- Scientific statements: tags are also means of communication for data collection issued from scientific statements (monitoring) or autonomous measuring sites (weather stations).
- Traceability: these subcutaneous radio-tags, originally designed for animal traceability, can be used for humans without any technical constraint.

The company Applied Digital Solutions is indeed proposing subcutaneous radio-tags (commercially called VeriChip) as a solution to identify frauds, ensure safe and protected access to confidential sites, to store medical data, etc. For instance, Mexico City implemented 160 radio-tags under the skin of their policemen to control their access to databases and for better localization in case of emergency.

- Supervision: combined with sensors sensitive to the principal functions of human body, RFID systems can also be an integrated solution for real-time supervision the health of a patient.

6 APPLICATIONS OF RFID

6.1 *Most existing applications*

Among existing applications, we can list:

- Access to public transportations as mentioned earlier,
- Identification and control of merchandise storage and inventory;
- Industrial tracking of assembly channels;
- Automatic data acquisition of a list of products bought or left stock;
- Easier access to commodities and goods (such as books in libraries);

6.2 *Other issues*

Smart tags are often considered like a means of replacing and improving the barcodes under the UPC/EAN standard.

The radio-identifiers IDs are indeed rather long and countable to give to each object a unique number, whereas UID codes used currently only make it possible to give a number for a class of products. This property of the RFID makes it possible to trace the displacement of objects from a place to another, from the chain of production to the ultimate consumer. This makes such technology as the most suitable key to logistic issues.

However, RFID systems, although they are operational, suffer from a lack of normalization.

The multitude of solutions suggested by various manufacturers makes the universal traceability difficult to realize.

EPC global (EPC) 2009 is working toward this aspect to propose an international standard in order to standardize the different technical uses of RFID. The goal is to be able to have a homogeneous international system of distribution of identifiers so that one can have an electronic EPC product code for each object present in the logistic chain of a company worldwide.

The properties of the radio-tags would also make it possible to consider applications bound for the ultimate consumer such as:

- Appliances (like refrigerators) capable not only to automatically recognize and provide all products they contain, but also to control the optimal date for a given product to be safely consumed;
- Clothing (storage and inventory);
- Automatic identification of postal codes;
- Traceability of new-born babies.

7 ENVIRONMENTAL IMPACT

Like all industrial products, RFID chips consume natural resources and produce no recycle items.

It is unfortunate that only very few studies have been conducted to investigate the direct environmental impact of this technology (AFSSET) 2008. However, the RFID industry is making great developments in this direction, in particular to answer the environmental stakes in sensitive areas like production chains, waste management, transportation and geo-localization.

For example, in certain European cities, the residential dustbins are equipped with RFID chips. The trucks dustbins, equipped with RFID readers, identify the dustbins collected thanks to their RFID chips. This management of waste by RFID allows a better monitoring of their nature and their quantity in order to optimize their treatment.

8 SECURITY ISSUES IN RFID SYSTEMS

The RFID Technologies could appear dangerous for the human and the society (health and private life protection) (Bellaire) 2005, with:

- Possibility to access to sensitive information for both private entities/companies and/or governmental structures;
- Use of information contained in tags implemented in passports to selectively target persons or group of persons;

- Abusive use of databases of people having bought or borrowed certain types of sensitive goods (weapons, ...);
- Potential problems of “numerical/economical sovereignty” related to the infrastructure of the EPCGlobal network;
- Implantation of subcutaneous chips in terms of ethical issues and rights to the physical integrity of a person. Under certain conditions, people refusing these subcutaneous tags could likely be victims of discrimination;
- Identification of people by their signature (bank cards, mobile phones, public transportation passes...);
- Identification and localization of persons using RFID tagged objects;
- Beyond a certain threshold, emission of RF signals has been proved to be dangerous for health, in particular the multiplication of cancers in the case of experiments on mice or interferences that can disturb the operation of the biomedical apparatuses (van der Togt) 2008.
- In a report published on 2009 (AFSSET) 2009, AFSSET recommended to continue the development of scientific search for biological effects of the radiations related to RFID.

To protect their citizen privacy, some country legislations provide a certain protection on citizen private life by forbidding any hidden control or identification. Also, the use of the same apparatuses for the access control and the control of presence (FoeBuD) 2010, (Liberation/écrans, 2006), and (IPC, 2004).

In 2006, a group of hackers announced in the Bi-annual HOPE convention, that they have successfully cracked the safety features of the subcutaneous chip (VeriChip's) 2006. They also claim to have been able to reproduce it. They estimate that the legislation is too flexible with this technology, taking into consideration its potential risk for private life protection.

9 CONCLUSION

Nowadays, the use of RFID has become a habit in people's daily life, in many sectors such as access control, merchandise inventory, asset tracking, traceability of objects and animals, identification of readers in libraries, to name a few.

However, the rapid development of this technology opens the door to unauthorized and/or criminal data manipulations, with clear implications in terms of privacy and data.

Another issue is the potential impacts of this technology on our environment and health.

REFERENCES

- AFSSET. 2008. French Agency for the Safety of Environment and Labour, p. 98.
- AFSSET. 2009. French Agency for the Safety of Environment and Labour.
- Bellaire A. 2005. Dossier futura-sciences. Puce RFID: mythes and realities of the Big Brother miniaturized.
- Capurro, R. 2010, Ethical Aspects of ICT Implants in the Human Body presentation. IEEE International Symposium on Technology and Society. New South Wales: IEEE.
- COE: <http://conventions.coe.int/treaty/fr/treaties/html/164.htm>. Convention of European council on Human Rights and Biomedicine], signed on april, 4th, 1997 at Oviedo (see specially article 5 to 10).
- Convention of the Council of Europe, January, 1st 1981 for the Protection of Individuals with regard to Automatic Processing of Personal Data.
- EC: <https://ec.europa.eu/research/eye/index.cfm>, 2009.
- Opinion of the French Agency for the Safety of Environment and Labour—AFSSET; epcglobalinc.org. European Recommendation, May, 12th 2009.
- GEESNT Groupe européen d'éthique des sciences et des nouvelles technologies, 2005. Aspects éthiques des implants TIC dans le corps humain. *Avis du groupe européen d'éthique des sciences et des nouvelles technologies*, PDF, 39 pages, consulted 2016-01-04.
- Finkenzeller, K. 2003. *RFID Handbook*, 2nd Ed. The Atrium, England: J. Wiley & Sons.
- FoeBuD. 2010. German Association FoeBuD to prevent potential abuse radiolabels.
- Hitachi, 2006. World's smallest and thinnest 0.15 x 0.15 mm, 7.5µm thick RFID IC chip—Enhanced productivity enabled by 1/4 surface area, 1/8th thickness, pdf. Tokyo: Hitachi, consulted 2016-01-04.
- IPC: <https://www.ipc.on.ca/images/resources/up-rfid.pdf>, 2004.
- JO, France, 201 du 31.7.2002, p. 37-47.
- Le RFID au service d'une gestion rationnelle des déchets, *Greenit.fr*, 2014.
- Lee S.R. Joo S.D and Lee C. W. 2005. An Enhanced Dynamic Framed Slotted Aloha Algorithm for RFID Tag Identification. *In the 2nd International Annual Conference on Mobile and Ubiquitous Systems: Networking and Services, San Diego, 2005*. San Diego: Springer.
- Liberation/écrans—Interview of Mélanie Rieback, 2006.
- Myung, J., Lee W. and Srivastava J.2006. Adaptive Binary Splitting for Efficient RFID Tag Anti-Collision. *IEEE Communication Letters*, vol. 10, no. 3: 144-146.
- Nsanze, F. 2005. Rapport, ICT implants in the human body—A Review. *The European Group on Ethics in Science and New Technologies to the European Commission*: 115-154.
- Perret, E. 2014. *Identification par radiofréquence*, Grenoble: ISTE Edition.
- Point 58 (NTIC) and point 59 (usages abusif des TIC) the Declaration of the World Summit principles on the Information Society (2003-12-12) on the use of Information and Communications Technology (ICT). Rective 90/385/CEE of Council of june, 20th, 1990. On the approximation of the laws of Member States

- relating to active implantable medical devices (JO L 189 du 20.7.1990, p. 17–36).
- Saadi H. Touhami R. Yagoub M.C.E. 2011. Simulation of the anti-collision process of RFID system based on multiple access protocols modelling. *IEEE Int. Symposium on Signal Processing and Information Technology*, Bilbao, Spain, 2011.
- UNESCO: http://portal.unesco.org/shs/fr/ev.php-URL_ID=2228 & RL_DO=DO_TOPIC&URL_SECTION=201.html. Universal Declaration on the Human Genome and Human Rights], adopted by the UNESCO, November, 11th 1997.
- Van Der Togt R, Jan Van Lieshout E, Hensbroek R, Beinat E, Binnekade JM, Bakker PJM. 2008. Electromagnetic interference from radio frequency identification inducing potentially hazardous incidents in critical care medical equipment, *JAMA*, 2008; 299: 2884–2890.
- VeriChip's. 2006. Human-implantable RFID chips cloneable, sez hackers. Engadget 24/07/2006.
- Vogt H. 2002. Efficient Object Identification with Passive RFID Tags, in *Pervasive Computing: 1st International Conference*, vol. 2414/2002. Switzerland: Springer Berlin/Heidelberg: 98.
- Yu J. Liu K. Huang X. Yann G. 2008. A Novel RFID Anti-collision Algorithm based on SDMA. *Wireless Communications, Networking and Mobile Computing Int. Conf.*, 2008.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

The impact of using educational gamification in mobile computing course: A case study

Rula Al-Azawi & Mohaned Al-Obaidy

Department of Computer Science, Gulf College, Seab, Oman

Aladdin Ayesh & Duska Rosenberg

Faculty of Technology, De-Montfort University, Leicester, UK

ABSTRACT: The purpose of this case study was to explore a new learning strategy that benefits the students of mobile computing modules to engage and understand the mobile computing course through designing computer games which enhance their understandability concept.

Games can be used within higher education in various ways to promote student participation, enable variation in how lectures are taught, and improve student interest.

The game applied to be designed by students focusing also to design educational game for children with age six years. This game will teach the students how to learn in a fun way.

Our case study is implemented at Gulf College which is affiliated with Staffordshire University-UK. Our game design was applied to teach students Android Studio software by designing an educational game. We describe the important requirements need to be considered when students start their design and implementation process.

Finally, we describe the findings and results of our case study. The data analysis and evaluation are based on students feedback, staff feedback and the final marking grades for the students.

1 INTRODUCTION

The traditional teaching method is a book, a piece of chalk and black board. Such teaching method currently has not suitable to our students and especially to science and technology students. We need to new means of information transfer more suitable to our students. It is important to use new teaching tools and methods to improve teaching effectiveness.

In this paper, we have used new teaching methods based on game design to engage students and get their interest to work perfectly with their assignment. The main topic of this work is development of educational games that can be used on mobile devices. Paper describes a novel approach to educational game defining and interpretation. Idea is based on extracting knowledge, game rules and scenarios outside the program thus enabling reusability. Lack of creativity and innovation in schools has also been attributed to technology design. Technologies are often designed for the market rather than for education (Milo).

The human ability to realize graphic representations faster than textual representations led to the idea of using graphical artifacts to describe the behavior of algorithms to learners and animations

to communicate (Shabanah, Chen, Wechsler, Carr, & Wegman 2010).

Purpose of this is to take learning outside classrooms and homes and provide a fun and interesting way of learning anytime, anywhere (Milo).

This study tries to enlarge the scope of gameplay in learning situation by adopting gamified learning strategy and combining game elements with well-designed mobile learning activities. The goal is to determine whether mobile technologies can support gamified learning approach, and learning strategies influenced achievement in the natural science course. The implementation of our solution proposal, called Mobile Insect Learning System (MILS). A series of learning activities geared toward integrating the game elements into course design was developed for the outdoor learning environment (Su & Cheng 2013).

The rest of paper is organized as follows. Section two describes the related work of combination between education concept and game design. Section three describes the usage of computer game in education in both directions, first as gamification in education and second as Game Based Learning (GBL) (Kalloo 2010). Section four describes our case study of applying designing of educational game as a part of mobile computing

course. Section five presents the finding of our case study which was based on students feedback, staff feedback and final mark grade of the course. Finally, section six presents the conclusion and future work.

2 RELATED WORK

The use of games in the educational context has had a positive impact on motivation and other skills (Wastiau 2009). This study shows that students were generally happy that teachers integrated applications from their everyday reality into their educational process. Teachers, on the other hand, found that the use of games in teaching improved students' self-confidence and also were more appropriate when it came to mistakes and different learning rhythms of students (Anusca Ferrari & Punie 2009).

Preliminary research suggests that mobile devices can create more active learning experiences that improve student engagement, learning, and course retention (Joosten 2010), and the use of new technologies can enhance motivation, which is a vital aspect of learning, deliver information when needed, and encourage to solve problems and satisfy curiosity. Furthermore, much research has shown that the mobile device as a mobile guide can help student to increase their science and geographical knowledge as well as their motivation to engage in learning activities (Huang 2010).

Mobile technologies can meet higher order learning needs and realizing a more creative and learner-centered educational process (Joo 2009).

Several studies have demonstrated successful experiments that support knowledge production and transmission among learners and educators through the use of mobile devices in the learning activities of various courses (Ogata 2009) (Sandberg 2011) (Su & Cheng 2013).

Both studies for (Wang & Wu 2011) (Wang & Wu 2009) focuses in the work to show how the technical aspect such as games can be applied in computing courses.

Games can mainly be integrated into higher education in three ways. First, traditional exercises can be replaced by games motivating the students to put extra effort in doing the exercises, and giving the course staff an opportunity to monitor how the students work with the exercises in real-time (Sindre, Natvig, & Jahre 2009) (Foss & Eikaas 2006). Second, games can be used within a traditional classroom lecture to improve the participation and motivation of the students through knowledge-based multi-player games played by the students and the teacher (Wang, March-Storstein, & Fsdahl 2007) (Wang, Fsdahl, & Morch-Storstein 2008).

Third, game development projects can be used in Computer Science (CS) or Software Engineering (SE) courses to learn specific CS or SE skills (El-Nasr & Smith 2006) (Wu & Wang 2009).

3 COMPUTER GAME IN EDUCATION

In this section, we will explain innovative use of computing technology to assist students to design their games using Android Studio as a higher education institution. Although educational games are accepted in elementary school, teacher and parent interest in their use declines in the later grades (Gredler).

In today's information society, digital learning has the features of not being constrained by time and space, being more attractive to learning attention of students compared to traditional instruction, can better increase learning motivation, promote problem-solving ability, and in turn achieving better learning effects (Chen, Jian, Lin, Yang, & Chang 2014).

Research on games has also demonstrated that when games are used in educational contexts, appropriation can take place on different levels. In educational contexts, learners need to be able to enter the world of the game, but also be critical about the process, so as to be able to reflect upon their relationship with the game when viewed from outside. This suggests that creative learning through gaming requires substantial effort from teachers, in order to achieve positive results. Research carried out by the European Schoolnet demonstrates that when teachers used games in their teaching, pupils' motivation and skills were increased (Wastiau 2009) (Anusca Ferrari & Punie 2009).

The success of research-based educational games, when presented commercially, led to the production of commercial educational games. Usually, educational games are known as Edutainment games for embedding education with entertainment.

However, despite the use of educational computer games in teaching many subjects, still there is a need for more games to teach several other subjects.

In the next section, we will explain two methods of using game in education.

3.1 Gamification in education

Before we start our section, it is important to understand **What is gamification?** Gamification originates from the computer games industry and is the use of game thinking and game mechanics in a non-game context in order to engage users, solve problems and drive behavior.

On a basic level gamification techniques tap into and influence peoples natural desires for competition, achievement, recognition and self-expression. Gamification appears to be making the leap from game-play to the workplace at a great pace. A growing number of organizations are adopting gaming techniques and game-style rewards in order to motivate and incentivize employees and customers (Gartner 2011).

Reviewing education and the courses establishments offer is a good indicator of how gamification is viewed and being utilized in business. Several UK and US universities offer courses encompassing gamification, ranging from one-off courses to a module on a masters degree. Put at its simplest, gamification is the use of game design elements, game thinking and game mechanics to enhance non-game contexts. This is the main function that gamification could provide enhancing a situation through the use of gaming mechanics, the benefits of which include (APM Thames Valley 2014):

- Increased engagement
- Higher motivation levels
- Increased interaction with the user (customer or employee)
- Greater loyalty

Finally, Gamification Mobile technologies can be used as powerful cognitive tools within constructivist approach to solve complex problems and to engage students in authentic and meaningful activities (Jonassen Reeves, 1996).

Educational gamification proposes the use of game-like rule systems, player experiences and cultural roles to shape learners behavior. (Sandberg 2011) found that many children used a trial-and-error strategy on play the games. For this reason, gamifying a course would be a great help to primary students by take advantage of the motivational power of games and apply it to the motivational problems in education so successful learning can take place (Su & Cheng 2013).

3.2 *Game Based Learning (GBL)*

In 2006, there has been rapid growth of studies relating to digital game-based learning, which shows that this research issue is occupying a more important position in the education field. Research of many scholars has also pointed out that learning based on digital games can increase the learning interest and motivation of students.

Developers and researchers are working in various areas of game-based learning, including games that are goal-oriented; social game environments; non-digital games that are easy to construct and play; games developed expressly for education; and commercial games that lend them-

selves to refining team and group skills (Singer & Johnsson 2015).

Digital Game-based Learning Ideal digital game-based learning makes people feels as if they are playing videogames or computer games. The operational value of digital games is in the educational benefit for student development and learning. Even games that do not deliberately incorporate learning elements can help the cognitive development of students. In the learning process, digital games have the two important learning elements of being interesting and fun, thus digital games can help learners to be in an effective learning environment that is at ease and with stronger learning motivation, so that learners can use digital game-based learning to develop the basic techniques and knowledge in specific fields necessary in the digital technology age.

Children also believe that digital game—based learning helps them to learn faster, and have greater interest in focusing on learning topics. Most students feel that mathematics is a difficult subject, and many students lose their learning motivation in response to the repetitive and monotonous mathematical learning in the classroom, having lost morale for learning mathematics (Kalloo 2010). If it is possible to use digital game-based learning for mathematics, students not only think that mathematics has become more interesting, but teachers and parents also think that if students use games to learn mathematics, it can effectively enhance their mathematical knowledge and abilities (Chen, Jian, Lin, Yang, & Chang 2014).

For this reason our educational game will based on learning math and English in a funny and easy way at the same time.

4 CASE STUDY: APPLYING GAME CREATION IN MOBILE APPLICATION COURSE

Nowadays, mobile devices could provide us a great opportunity to engage students to learn outside the classroom and in a fun way. This section introduce the research concept of our case study which applies the mobile game with game elements to facilitate the mobile activity in an educational environment.

We designed an experiment in which students carry out the well-designed gamified learning activity in a mobile learning environment. To evaluate the effectiveness of this approach, Quantitative data analysis was also used to evaluate the students Learning achievement, staff feedback and students feedback. The study proposes a research framework relative to the effectiveness of learning achievement by adoption of different learning

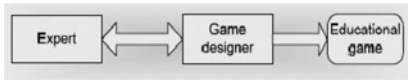


Figure 1. Mobile learning environment.

approach in a mobile learning environment as shown in Figure 1 (Su & Cheng 2013).

4.1 Game creation and prototype

When educational computer games were first introduced for widespread use, many rudimentary games were little more than some form of academic drill and practice (e.g., multiplication problems, spelling exercises), with correct answers rewarded by an animation of, e.g., clowns or fireworks. Thus, the actual game play was not terribly different than doing a series of pencil-and-paper arithmetic problems in school, although the payoffs were more entertaining. Indeed, while technology and graphics may be far more advanced today, the same type of underlying approach nevertheless still appears occasionally today (e.g., in electronic toys that present children with decontextualized arithmetic problems and reward correct answers with music or sound effects) (Fisch) (Navarrete 2013).

Game creation is suggested to provide greater student engagement and learning through game play (Vos et al., 2011). The purpose of this case study is to examine the learners creative thinking experience in the digital game design and development course with primary-school students.

Using of Android Studio as a mobile programming language provides an interesting and useful context in which continue searching and increase their independent study time.

In our game prototype, we advice our students to focus on the following points in their game creation process:

- Supporting game play.
- Enhance Learning via feedback and hints.
- Feedback for the wrong answer.
- Help section

Furthermore, students should follow game design elements below:

- Game idea: Describes the game main goal and topic.
- Game start: Describes the game start up screen components.
- Game level: Describes how the difficulty increases, how a level ends. Each completed level must achieve a learning sub-goal.
- Game end: Explains what happens when the player loses, or wins or gets a high score.

- Game graphics: The game graphics must assist child to play the game in a fun way.
- Game play: explains how the game is played.
- Game sounds: musical sounds and some of animals sounds should be added to the game.

4.2 Students game

The use of mobile learning is a trend in higher education and is redefining the manner in which learning takes place and how instruction is delivered (Geist 2011). Mobile learning presents students and professionals with the unique opportunity to access information instantaneously regardless of location. This means that learning can occur anywhere at any time through the use of these devices.

In the classroom, the use of mobile devices has been found to contribute to the learning experience and engage students during lectures. Students perceive the tablet PC to be effective in improving their learning environment. Moreover, students report the tablet PC to facilitate their ability to understand key concepts and personalize their learning experience. The use of tablet computers helped to create a cooperative learning environment among students (Foti, Drive, & Ave 2014).

The Figure 3 and Figure 4 shows sample of students educational games. Some students based to reflect directly to player if they select wrong answer. Other group of students based to provide final score at the end of the game in Figure 2. Finally, all of our students games add some sounds and music to enhance children knowledge to recognize letters, numbers and animal sounds.

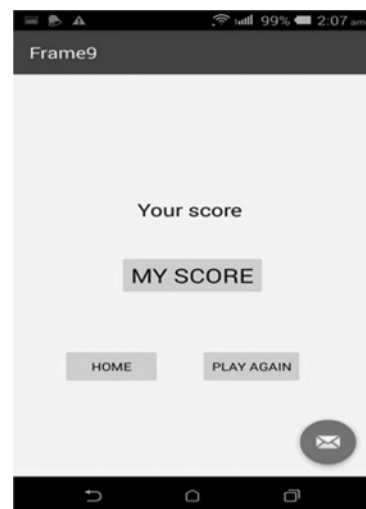


Figure 2. Students sample works.



Figure 3. Students sample works.



Figure 4. Students sample works.

5 FINDINGS

In this study, we have measured the students engagement in the classroom and the effect of different teaching strategy on learning and compare the result with previous students results where the students were implementing no game assignment. There are totally 37 students participate in this study of which 20 are female and 17 are male with age range 20–40 years. This is because we delivered part time and full time study mode.

5.1 Staff feedback and class observation

This case study was run with two academic staff members. Both of us agree that students increase experience with domain of creative process. The students learnt programming that using creative thinking in problem solving and learning activities.

The following points reflect the staff indication at the end of the course:

- Increase the creative thinking and personal perception
- Deal positively with technological challenge.
- Increase the independent study time.
- Deep level of thinking and enhance the ability of problem solving.
- Trying to do the best to satisfy the customer even if they are just children.

5.2 Students results

This section based on the results grade sheet of our case study. We also compare their results with pre-

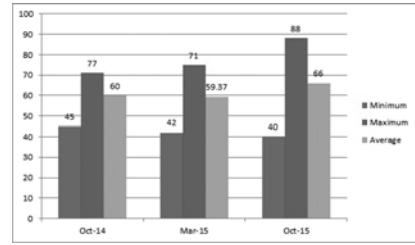


Figure 5. Comparative Students results for last three semesters.

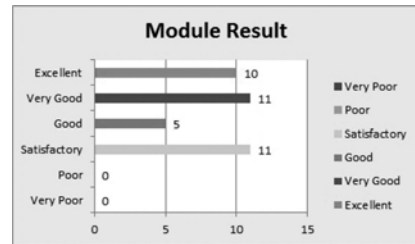


Figure 6. Students results.

vious group of students when they implemented non game assignment for this course as shown in Figure 5. In Figure 5, we have notice that one of our students gets 88 mark in his final result and the average of students marks is 66 which is greater than previous semesters.

We have notice that the average marks in this semester has been increased as shown in Figure 6 and we have full students engagement in the classroom. The impressive results was no fail student during our case study. This is because students feel challenged and fun while doing their assignment.

5.3 Students feedback

As a part of the college evaluation system, the students should fill a feedback form before end of the course. The feedback for this semester was positive and students enjoy doing their assignment. Furthermore, our students fill a questionnaire related to our case study. Most of the student mention that they would like to join to game group and to design a game as a final year project for educational or simulation game. Also, they have spend extra time searching on web site and through YouTube to add extra features to their game design.

6 CONCLUSION AND FUTURE WORK

This paper aimed to investigate how we could use gamification in education approach through mobile learning application and how it effects on

students engagements and learning outcome. Our experiment design was based to let students create an educational game as their assignment part and compare the result with previous group of students where their assignment was non game application. We have notice that creating an effective educational game entails much more than simply creating an engaging game and building in age-appropriate educational content.

Through gamification we can not only create a mindset that encourages students to try new things, to not be afraid of failing, but also can enable students to engage in enjoyable experiences for the purpose of learning. In addition, gamification is an innovative approach to learning, and because new technologies and new applications are continuously emerging, it is still developing. Future studies must continue to examine the new mechanics and new applications associated with emerging gamification technologies (Su & Cheng 2013).

As a future work, most of our students in this case study are willing to create a game as a part of final year project process. As an academic member, we have plan to implement our innovation teaching method to be a part of extra computing courses.

REFERENCES

Anusca Ferrari, R.C. & Y. Punie (2009). Innovation and Creativity in Education and Training in the EU Member States: Fostering Creative Learning and Supporting Innovative Teaching. Technical report, European Communities.

APM Thames Valley (2014). *Introduction to Gamification*.

Chen, H.R., C.H. Jian, W.S. Lin, P.C. Yang, & H.Y. Chang (2014). Design of Digital Game-Based Learning in Elementary School Mathematics. *2014 7th International Conference on Ubi-Media Computing and Workshops*, 322–325.

El-Nasr, M.S. & B.K. Smith (2006). Learning through game modding. *Computers in Entertainment* 4(1), 45–64.

Fisch, S.M. Making Educational Computer Games Educational? MEDIA CONTENT. (1).

Foss, B.A. & T.I. Eikaas (2006). Game play in engineering education concept and experimental results. *International Journal of Engineering Education* 22(5), 1043–1052.

Foti, M.K., V.K. Drive, & M. Ave (2014). Mobile Learning: How Students Use Mobile Devices to Support Learning. *15*(3).

Gartner (2011). *Gartner says by 2015, more than 50 per cent of organisations that manage innovation processes will gamify those processes*. press release.

Geist, E. (2011). The game changer: using ipads in college teacher education classes. *College Student Journal* 45(4), 758–768.

Gredler, M.E. Games and Simulations and their Relationships to Learning. (d), 571–582.

Huang, Y.M., L.Y.T. C.S.C. (2010). Effectiveness of a mobile plant learning system in a science curriculum in taiwanese elementary education. *Computers and Education* 54(1), 47–58.

Joo, K.H., K.S.H. (2009). Development and application of an efficient ubiquitous teaching and learning model. In *InICACT 2009, 11th International Conference on Advanced Communication Technology*, pp. 2165–2168.

Joosten, T. (2010, June 9–12). Mobile learning and social media: Increasing engagement and interactivity. anaheim, ca. In *New Media Consortium*, Anaheim, CA.

Kaloo, V., K.. M.P. (2010). Personalized game based mobile learning to assist high school students with mathematics. In *IEEE International Conference on Advanced Learning Technologies*, pp. 485–487.

Milo, M. Mobile educational game: adventure anywhere.

Navarrete, C.C. (2013, nov). Creative thinking in digital game design and development: A case study. *Computers and Education* 69, 320–331.

Ogata, H., M.YE.-B.M.M. Y.Y. (2009). Lorams: linking physical objects and videos for capturing and sharing learning experiences towards ubiquitous learning. *International Journal of Mobile Learning and Organisation* 3(4), 337–350.

Sandberg, J., M.M. d. G.K. (2011). Mobile english learning: An evidence-based study with fifth graders. *Computers & Education* 57(1), 1334–1347.

Shabanah, S.S., J.X. Chen, H. Wechsler, D. Carr, & E. Wegman (2010). Designing Computer Games to Teach Algorithms. *2010 Seventh International Conference on Information Technology: New Generations*, 1119–1126.

Sindre, G., L. Natvig, & M. Jahre (2009). Experimental validation of the learning effect for a pedagogical game on computer fundamentals. *IEEE Transactions on Education* 52(1), 10–18.

Singer, K. & C. Johnsson (2015). A Game—Based Method for Teaching Entrepreneurship. (1), 51–65.

Su, C.H. & C.H. Cheng (2013, nov). A Mobile Game-based Insect Learning System for Improving the Learning Achievements. *Procedia—Social and Behavioral Sciences* 103, 42–50.

Wang, A., T. Fsdahl, & O. Morch-Storstein (2008). An evaluation of a mobile game concept for lectures. In *21st Conference on Software Engineering Education and Training (CSEET 08)*, Number 197–204.

Wang, A., O. March-Storstein, & T. Fsdahl (2007, November). Lecture quiz mobile game concept for lectures. In *11th IASTED International Conference on Software Engineering and Application (SEA 07)*.

Wang, A.I. & B. Wu (2009). An Application of a Game Development Framework in Higher Education. *International Journal of Computer Games Technology* 2009, 1–12.

Wang, A.I. & B. Wu (2011). Using Game Development to Teach Software Architecture. *International Journal of Computer Games Technology* 2011, 1–12.

Wastiau, P., K.C. V.W. (2009). How are digital games used in schools? Technical report, European Schoolnet.

Wu, B. & A.I. Wang (2009, February). An evaluation of using a game development framework in higher education. In *22nd Conference on Software Engineering Education and Training (CSEET 09)*, Hyderabad, India, pp. 41–44.

Efficient mining of high average-utility itemsets

Jerry Chun-Wei Lin & Ting Li

School of Computer Science and Technology, Harbin Institute of Technology Shenzhen Graduate School, Shenzhen, China

Philippe Fournier-Viger

School of Natural Sciences and Humanities, Harbin Institute of Technology Shenzhen Graduate School, Shenzhen, China

Tzung-Pei Hong

Department of Computer Science and Engineering, National University of Kaohsiung, Kaohsiung, Taiwan

Miroslav Voznak

Department of Telecommunications, VSB-Technical University of Ostrava, Ostrava, Czech Republic

ABSTRACT: In traditional High-Utility Itemset Mining (HUIM), the utility of an itemset is defined as the sum of the utilities of its items in transactions where it appears. An important problem with this definition is that it does not take itemset length into account. To provide a better assessment of each itemset's utility, the task of High Average-Utility Itemset Mining (HAUIM) was proposed and several algorithms have been extensively studied. Most of the past works are based on level-wise or pattern-growth approaches, which required amounts of computation to mine the required High Average-Utility Itemsets (HAUIs). In this paper, we present an efficient Average-Utility (AU)-list structure to discover the HAUIs more efficiently. A depth-first search algorithm named HAUI-Miner is proposed to explore the search space without candidate generation, and an efficient pruning strategy is developed to reduce the search space and speed up the mining process. Extensive experiments are conducted to compare the performance of HAUI-Miner with the state-of-the-art algorithms of HAUIM in terms of runtime and number of determining nodes.

1 INTRODUCTION

Mining Frequent Itemsets (FIs) or Association Rules (ARs) in transactional databases is a fundamental task in Knowledge Discovery in Databases (KDD) (Agrawal & Srikant 1994, Agrawal, Imielinski, & Swami 2005). The most common ways of deriving FIs or ARs from a database are to use a level-wise (Agrawal & Srikant 1994) or a pattern-growth approach (Han, Pei, Yin, & Mao 2004). Traditional algorithms of FIM or ARM only consider, however, occurrence frequencies of items in binary databases. Other important factors such as quantities, profits, and weights of items are not taken into account by traditional FIM and ARM algorithms. Thus, high-utility Itemset Mining (HUIM) has emerged as a critical issue in recent decades, as it can reveal the profitable itemsets in real-world situations (Liu & Qu 2012, Liu, Liao, & Choudhary 2005, Yao, Hamilton, & Butz 2004). HUIM can be considered as an extension of FIM that considers additional information such

as quantities and unit profits of items, to better assess how “useful” an itemset is to the user. Several level-wise and pattern-growth algorithms have been proposed to efficiently mine HUIs (Ahmed, Tanbeerand, Jeong, & Lee 2009, Chan, Yang, & Shen 2003).

In traditional HUIM, the utility of an itemset is defined as the sum of its utilities in the database. An important problem with this definition is that it does not take itemset length into account. To provide a better assessment of each itemset's utility, the task of High Average-Utility Itemset Mining (HAUIM) was proposed by Hong et al. (Hong, Lee, & Wang 2009). This measure addresses the bias of traditional HUIM toward larger itemsets, by considering the length of itemsets, and can thus more objectively assess the utility of itemsets. As for traditional HUIM, several algorithms have been designed for HAUIM (Hong, Lee, & Wang 2009, Lan, Hong, & Tseng 2012, Lin, Hong, & Lu 2010). In this paper, we first design an efficient Average-Utility (AU)-list structure and develop

an HAU-Miner algorithm for mining the HAUIs without candidate generation. The key contributions of this paper are threefold.

1. We design an efficient HAU-Miner algorithm to mine the High Average-Utility Itemsets (HAUIs) based on the developed Average-Utility (AU)-list structure. This structure can only keep the necessary information for later mining process through simply operation of intersection.
2. An efficient pruning is developed to reduce the search space of the enumeration tree and the unpromising candidates can be early pruned, thus improving the performance of runtime and reducing the number of determined nodes in the search space.
3. Substantial experiments are conducted to evaluate the performance of the designed HAU-Miner algorithm in terms of runtime and the number of determined nodes compared to the state-of-the-art algorithms.

2 RELATED WORK

High-Utility Itemset Mining (HUIM) (Liu & Qu 2012, Liu, Liao, & Choudhary 2005, Yao, Hamilton, & Butz 2004), an extension of frequent itemset mining, is based on the measurement of internal utility and external utility. The purpose of HUIM is to discover the complete set of High-Utility Itemsets (HUIs), that is itemsets having a utility no less than a minimum utility threshold. Yao et al. proposed a framework for mining HUIs based on mathematical properties of the utility measure (Yao, Hamilton, & Butz 2004). Since the Downward Closure (DC) property of ARM does not hold in traditional HUIM, Liu et al. then designed a Transaction-Weighted Downward Closure (TWDC) property and developed the Transaction-Weighted Utilization (TWU) model (Liu, Liao, & Choudhary 2005). However, the TWU model still requires to generate numerous candidates to obtain the actual HUIs. Pattern-growth algorithms have been proposed to compress the database into a condense tree structure using the TWU model. Lin et al. designed a High-Utility Pattern (HUP)-tree algorithm to recursively mine high-utility itemsets (Lin, Hong, & Lu 2011) using the proposed tree structure. Tseng et al. developed the UP-Growth+ (Tseng, Shie, Wu, & Yu 2013) algorithm to efficiently discover HUIs based on different pruning strategies. As an alternative to the pattern-growth mechanism, Liu et al. developed the list-based HUI-Miner algorithm (Liu & Qu 2012) to discover HUIs without candidate generation. Several extensions of the task of HUIM have been proposed

such as discovering up-to-date HUIs (Lin, Hong, & Lu 2015) and top- k HUIs (Zihayat & An 2014).

In traditional HUIM, the utility of an itemset is defined as the sum of the utility of its items, in transactions where it appears, regardless of its length. Thus, the utility of itemsets tends to increase with their length, and this definition is thus inadequate in real-life situations. To better assess the utility of itemsets, the task of High Average-Utility Itemset Mining (HAUIM) was proposed (Hong, Lee, & Wang 2009). It consists of discovering the High Average-Utility Itemsets (HAUIs) in a transactional database. The average-utility of an itemset is defined as the sum of the utilities of its items, in transactions where it appears, divided by the number of items that it contains. This definition is thus more appropriate for real-world situations.

Similarly to traditional HUIM, several HAUIM algorithms have been designed using the TWU model. Lin et al. first developed the HAUP-tree structure and the HAUP-growth algorithm for mining HAUIs (Lin, Hong, & Lu 2010). Lan et al. proposed a Projection-based Average-utility Itemset mining (PAI) algorithm (Lan, Hong, & Tseng 2012) to reveal HAUIs using a level-wise approach. Lu et al. proposed the HAU-tree algorithm to further reduce the number of unpromising candidates for mining the actual HAUIs using a designed enumeration tree structure (Lu, Vo, Nguyen, & Hong 2014). However, mining HAUIs using the designed algorithm is still very time-consuming since the upper-bounds used by these algorithms are loose, thus numerous unpromising candidates need to be generated, and the recursive process for building the complete enumeration tree remains costly. In this paper, we first design an efficient Average-Utility (AU)-list structure and develop an HAU-Miner algorithm for mining the HAUIs.

3 PRELIMINARIES AND PROBLEM STATEMENT

3.1 Preliminaries

Let $I = \{i_1, i_2, \dots, i_m\}$ be a finite set of m distinct items. A quantitative database is a set of transactions $D = \{T_1, T_2, \dots, T_n\}$, where each transaction $T_q \in D$ ($1 \leq q \leq n$) is a subset of I and has a unique identifier q , called its *TID*. Besides, each item i_j in a transaction T_q has a purchase quantity denoted as $q(i_j, T_q)$. A profit table PT indicates the unit profit value of each item in the database as $PT = \{pr(i_1), pr(i_2), \dots, pr(i_m)\}$, where profit values are positive integers. A set of k distinct items $X = \{i_1, i_2, \dots, i_k\}$ such that $X \subseteq I$ is said to be a k -itemset, where k is the length of the itemset. An itemset X is said to be contained in a transaction T_q if $X \subseteq T_q$.

A minimum average-utility threshold δ is set according to the user's preference (a positive integer). An example quantitative database is shown in Table 1, which will be used as running example for the rest of this paper. This database contains six transactions and six distinct items, denoted with letters from (A) to (F). The profit table indicates the unit profit of each item appearing in the database, and is shown in Table 2. In the running example, the minimum average-utility threshold is set to ($\delta=16\%$).

Definition 1 The average-utility of an item i_j in a transaction T_q is denoted as $au(i_j, T_q)$, and defined as:

$$au(i_j, T_q) = \frac{q(i_j, T_q) \times pr(i_j)}{1}, \quad (1)$$

where $q(i_j, T_q)$ is the quantity of i_j in T_q , and $pr(i_j)$ is the unit profit value of i_j .

Definition 2 The average-utility of a k -itemset X in a transaction T_q is denoted as $au(X, T_q)$, and defined as:

$$au(X, T_q) = \frac{\sum_{i_j \in X \wedge X \subseteq T_q} q(i_j, T_q) \times pr(i_j)}{\frac{|X|}{\sum_{i_j \in X \wedge X \subseteq T_q} q(i_j, T_q) \times pr(i_j)}}, \quad (1)$$

where k is the number of items in X .

Definition 3 The average-utility of an itemset X in D is denoted as $au(X)$, and is defined as:

$$au(X) = \sum_{X \subseteq T_q \wedge T_q \in D} au(X, T_q). \quad (2)$$

Definition 4 The transaction utility of a transaction T_q is denoted as $tu(T_q)$, and defined as:

$$tu(T_q) = \sum_{i_j \in T_q} u(i_j, T_q). \quad (3)$$

Definition 5 The total utility of a database D is denoted as TU , and defined as the sum of all transaction utilities, that is:

$$TU = \sum_{T_q \in D} tu(T_q). \quad (4)$$

From the given example, the total utility of Table 1 is calculated as $TU = 32 + 16 + 22 + 28 + 37 + 15 (= 150)$.

3.2 Problem statement

The problem of mining high-average utility itemsets is to discover the complete set of High Average-Utility Itemsets (HAUIs). An itemset X is an HAUI in a database D if its utility is no less than the minimum average-utility count, specified by the user. The set of HAUIs is thus formally defined as:

$$HAUIs \leftarrow \{X \mid au(X) \geq TU \times \delta\}. \quad (5)$$

4 PROPOSED HAUI-MINER ALGORITHM

In this paper, we design an Average-Utility (AU)-list structure to store the information needed by the mining process. Moreover, an algorithm named HAUI Miner is also developed to mine HAUIs more efficiently than previous works. Besides, to restore the downward closure property and effectively reduce the search space, this paper introduces a Transaction Maximum Utility Downward Closure (TMUDC) property. It allows to prune unpromising candidates early, and thus to reduce the search space to efficiently discover the actual HAUIs.

Definition 6 The transaction-maximum utility of a transaction T_q is denoted as $tmu(T_q)$, and defined as the maximum utility of items in a transaction T_q that is:

$$tmu(T_q) = \max(\{u(i_j) \mid i_j \in T_q\}). \quad (6)$$

Definition 7 The average-utility upper-bound of an itemset X is denoted as $auub(X)$, and defined as the sum of the transaction-maximum utilities of transactions containing X , that is:

Table 1. A quantitative database.

TID	Transaction (item, quantity)
1	A:1, B:6, C:3, D:3, F:6
2	B:2, C:3, E:2
3	A:2, C:1, D:2, E:1
4	A:1, B:9, C:3, D:2, F:2
5	A:3, B:9, C:3, D:1, E:1
6	C:4, D:1, E:1

Table 2. A profit table.

Item	Profit
A	5
B	1
C	2
D	3
E	4
F	1

$$auub(X) = \sum_{X \subseteq T_q \wedge T_q \in D} tmu(T_q). \quad (7)$$

Definition 8 An itemset X is called an high average utility upper-bound itemset (*HAUUBI*) if its average utility upper-bound is no less than the minimum average-utility count. The set of all *HAUUBI*s is defined as:

$$HAUUBI \leftarrow \{X \mid auub(X) \geq TU \times \delta\}. \quad (8)$$

Theorem 1 (TMUDC property) The transaction maximal utility upper-bound measure is downward closed. The *TMUDC* property holds for any *HAUUBI* itemsets.

By Theorem 1, $auub(X^{k-1}) \geq auub(X^k)$. Therefore, if X^k is a *HAUUBI*, any subset X^{k-1} of X^k is also a *HAUUBI*.

Theorem 2 (HAUUBI \subseteq HAUIs) The *TMUDC* property ensures that *HAUUBI* \subseteq *HAUIs*. Thus, if an itemset is not a *HAUUBI*, none of its supersets are *HAUIs*.

Thus, if an itemset X is not a *HAUUBI*, it is also not a *HAUI*. This property can be used to reduce the search space by pruning numerous unpromising candidates, which speeds up the mining process.

4.1 The revised and projected databases

The proposed *HAUI-Miner* algorithm scans the database twice to calculate tight upper bounds on the average-utilities of candidate itemsets. During the first database scan, the set of high average-utility upper-bound 1-itemsets (*1-HAUUBIs*) is discovered. This latter is needed to construct the *AU-lists* of 1-itemsets. During the second database scan, 1-itemsets that are deemed non-*HAUUBI* (according to the minimum average-utility count) are removed. In other words, for an itemset X , if $auub(X)$ is less than the minimum average-utility count ($\delta \times TU$), X is not a *HAUUBI*, and X can thus be removed from the database. The database obtained after removing all such items from a database D is called the *revised database* of D , and is denoted as D' . After the original database has been revised, the sub-database corresponding to each item in 1-*HAUUBIs* is then projected, forming a smaller database that is used for constructing the corresponding *AU-list*. If an item is not an *HAUUBI* in the sub-database, the item is removed from the sub-database. In this way, a projected database is smaller than the revised database, and can thus accelerate the construction of *AU-lists*. For example, consider item (B); transactions are projected, forming the sub-database of (B). Then, the $auub$ value of each item in the sub-database is

compared with the minimum average-utility count to determine if it satisfies the condition to be an *HAUUBI*. The result of this process for (B) is shown in Table 3, which is called the *projected sub-database* of (B).

4.2 The Average-Utility (AU)-list structure

A projected database that has been revised twice can then be used to efficiently construct the average-utility-list (*AU-list*) structure of each item/set. The *AU-list* of an item/set X is a list of elements such that there is an element representing each transaction T_q where $X \subseteq T_q$. An element consists of three fields as (1) The *tid* field indicates the TID q of T_q . (2) The u field indicates the utility of X in T_q , i.e., $u(X, T_q)$. (3) The mu field indicates the transaction-maximum utility of X in T_q , i.e., $tmu(X, T_q)$. The *AU-lists* constructed using the projected sub-database of (B), depicted in Table 3, are shown in Figure 1.

In Figure 1, the first element (1, 6, 9) in the constructed *AU-list* of (B) indicates that (B) appears in transaction T_1 , has a utility of 6 in that transaction, and that the transaction-maximum utility of (B) in that transaction is 9. If the sum of the $u(X)$ values of all elements in an *AU-list* is no less than the minimum average-utility count, it is directly output as an high average-utility itemset (*HAUI*). To construct *AU-lists* of k -itemset ($k \geq 2$), it is unnecessary to rescan the original database. They can be constructed by performing an intersection operation using *AU-lists* of smaller itemsets (by comparing TIDs in *AU-Lists*). The construc-

Table 3. The revised projected sub-database of (B).

TID	Items
1	$B:6, A:1, D:3, C:3$
2	$B:2, C:3$
3	$B:9, A:1, D:2, C:3$
4	$B:9, E:1, A:3, D:1, C:3$

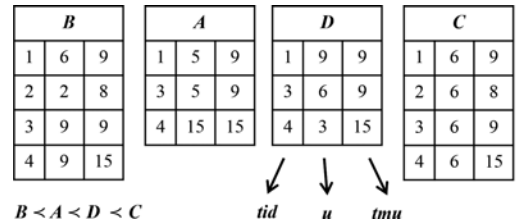


Figure 1. The *AU-lists* constructed using the projected sub-database of (B).

tion algorithm of AU-lists for k -itemsets ($k \geq 2$) is shown in Algorithm 1.

Algorithm 1: AU-list construction

Input: $P.AUL$, the AU-list of itemset P ;
 $P_x.AUL$, the AU-list of itemset P_x ;
 $P_y.AUL$, the AU-list of itemset P_y .
Output: $P_{xy}.AUL$, the AU-list of itemset P_{xy} .

```

1  $P_{xy}.AUL \leftarrow null$ ;
2  $P_{xy}.AUL \leftarrow null$ ;
3 for each element  $E_x \in P_x.AUL$  do
4   if  $\exists E_y \in P_y.AUL \wedge E_x.tid = E_y.tid$  then
5     if  $P.AUL = null$  then
6        $E_{xy} \leftarrow <$ 
7          $E_x.tid, E_x.u + E_y.u, E_y.tmu >$ ;
8     else
9       find  $E \in P.AUL$  such that
10       $E.tid = E_x.tid$ ;
11       $E_{xy} \leftarrow <$ 
12       $E_x.tid, E_x.u + E_y.u, E_y.tmu >$ ;
13       $P_{xy}.AUL \cup E_{xy}$ ;
14 return  $P_{xy}.AUL$ ;

```

4.3 Pruning strategy

Based on the designed AU-list structure, the search space for mining High Average-Utility Itemsets (HAUIs) can be represented as an enumeration tree, where each node represents a distinct itemset, which may be a potential HAUI. The proposed algorithm explores this tree using a depth-first search. To avoid this combinatorial explosion, this paper introduces an efficient pruning strategy that is effective at reducing the search space.

Definition 9 Let $SUM.X.iu$ denotes the sum of the utilities of an itemset X in a database D , that is:

$$SUM.X.u = \sum_{X \subseteq T_q \wedge T_q \in D} u(X, T_q). \quad (9)$$

Definition 10 Let $SUM.X.tmu$ denotes the sum of the transaction-maximum utilities of transactions containing an itemset X in a database D , that is:

$$SUM.X.tmu = \sum_{X \subseteq T_q \wedge T_q \in D} tmu(X, T_q). \quad (10)$$

Definition 11 Given an itemset X and a transaction T such that $X \subseteq T$, the set of all items appearing after X in T is denoted as T/X and defined as:

$$T/X = \{i \mid i \in T \wedge i \succ j \forall j \in X\}. \quad (11)$$

Definition 12 Let there be some itemsets X and Y . Y is said to be an extension of X if there exists an itemset $Z \neq \emptyset$ such that $Y = X \cup Z$, and

$\forall j \in Z, \exists i \in X$ such that $i \succ j$. Furthermore, Y is said to be a 1-extension of X if it is an extension of X and $|Z|=1$.

To mine HAUIs efficiently, it is necessary to reduce the search space. This can be done by identifying and pruning unpromising itemsets early. In the designed AU-list structure, the sum of the u and tmu fields provides enough information to achieve this goal.

Theorem 3 Let there be an itemset X . If the value $SUM.X.tmu$ calculated using the AU-list of X is less than the minimum average-utility count, all extensions of X are not High Average-Utility Itemsets (HAUIs).

Thus, if the sum of the transaction-maximum utilities of the transactions containing an itemset X is less than the minimum average-utility count, all extensions of X are not High Average-Utility Itemsets (HAUIs) and can thus be ignored, and their AU-lists do not need to be constructed. The full pseudocode of the proposed HAUI-Miner algorithm is presented in Algorithm 2.

Algorithm 2: HAUI-Miner

Input: $P.AUL$, the AU-list of itemset P ; $AULs$, the set of AU-list of all P 's 1-extensions;
 δ , the minimum average-utility threshold;
 TU , the total utility in D .

Output: all the high average-utility itemsets with P as the prefix.

```

1 for each AU-list  $Y.AUL$  in  $AULs$  do
2   if  $\frac{SUM.Y.u}{|Y.AUL|} \geq \delta \times TU$  then
3      $HAUIs \leftarrow HAUIs \cup Y$ ;
4   if  $SUM.Y.tmu \geq \delta \times TU$  then
5      $extAULs \leftarrow null$ ;
6     for each AU-list  $Z.AUL$  after  $Y.AUL$  in
7      $AULs$  do
8        $extAULs \leftarrow extAULs \cup$ 
9        $Construct(P.AUL, Y.AUL, Z.AUL)$ ;
10     $HAUI-Miner(Y.AUL, extAULs, \delta \times TU)$ ;

```

Table 4. Final derived HAUIs.

Itemset	au	Itemset	au
(B)	26	(BC)	25
(A)	35	(AD)	29.5
(D)	27	(AC)	27.5
(C)	34	(DC)	27.5
(BA)	24.5	(ADC)	26.3

The final set of HAUIs obtained for the running example is shown in Table 4.

5 EXPERIMENTAL RESULTS

In this section, the performance of the proposed HAUI-Miner algorithm is compared with the three state-of-the-art algorithms HAUP-growth (Lin, Hong, & Lu 2010), PAI (Lan, Hong, & Tseng 2012) and HAUI-tree (Lu, Vo, Nguyen, & Hong 2014) algorithms on several datasets. Experiments were conducted on three real-world datasets (Fournier-Viger & Lin) and one synthetic dataset generated using the IBM Quest Synthetic Data Generator (Agrawal & Srikant). A simulation model (Liu, Liao, & Choudhary 2005) was developed to generate quantities (internal utilities) and unit profit values (external utilities) of items in transactions for all datasets. External utilities have been generated in the $[0.01, 10]$ interval using a log-normal distribution, and internal utilities have been randomly chosen in the $[1, 5]$ interval. The characteristics of these datasets are shown in Table 5. In Table 5, $\#|D|$ is the total number of transactions; $\#|I|$ is the number of distinct items; **AvgLen** is the average transaction length; and **Type** is the dataset type.

In the performed experiments, if an algorithm ran for more than 10,000 seconds or if it ran out of memory, the algorithm was stopped.

5.1 Runtime

In this section, runtimes of the three state-of-the-art algorithms for mining HAUIs are compared with the proposed HAUI-Miner algorithm for various minimum average-utility threshold values, on the four datasets. Results are shown in Figure 2.

It can be observed in Figure 2 that the proposed HAUI-Miner algorithm outperforms previous algorithms for various minimum average-utility thresholds, on all four datasets. In particular, the proposed HAUI-Miner algorithm can be one to two orders of magnitude faster than the PAI, HAUP-growth and HAUI-tree algorithms. For example in Figure 2(b), the runtime of HAUP-growth, PAI, and HAUI-Tree are respectively 233.7, 4.5 and 6.6 seconds, while the proposed

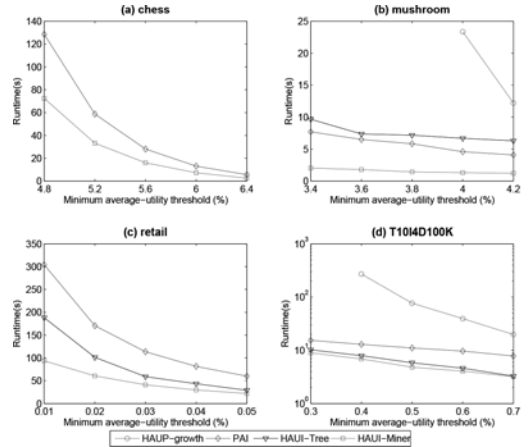


Figure 2. Runtimes w.r.t. variants of minimum average-utility thresholds.

algorithm only took 1.3 seconds when the minimum average-utility threshold was set to 4%. For the HAUP-growth algorithm, no results are provided in Figure 2(a) and Figure 2(c). Moreover, the HAUP-growth algorithm has no results in Figure 2(b) and 2(d) when the minimum average-utility threshold is respectively set to 3.8%, and 0.3% or below. The reason is that the HAUP-growth algorithm utilizes more memory to mine HAUIs based on its designed tree structure because it uses additional arrays to maintain the information to be used by the mining process as well as the HAUI-tree algorithm.

Another observation is that the gap between the designed HAUI-Miner algorithm and PAI is smaller when the minimum average-utility thresholds is set to large values, as well as with the other compared algorithms. This is also reasonable since when the minimum average-utility threshold is set higher, fewer candidates are generated and it is more easier to discover the actual HAUIs from a small set of candidates.

5.2 Node analysis

In this section, the number of nodes generated for discovering the actual HAUIs using each algorithm is compared. Results are shown in Figure 3.

It can be observed in Figure 3 that the number of nodes generated by the proposed HAUI-Miner algorithm is much less than for the HAUP-growth, PAI and HAUI-Tree algorithms for various minimum average-utility threshold values on all datasets. This is because the compared algorithms are all sensitive to transaction length. This is especially the case for the HAUP-growth algorithm since

Table 5. Characteristics of the datasets.

Dataset	$\# D $	$\# I $	AvgLen	Type
chess	3,196	75	37	Dense
mushroom	8,124	119	23	Dense
retail	88,162	16,470	10.3	Sparse
T10I4D100 K	100,000	870	10.1	Sparse

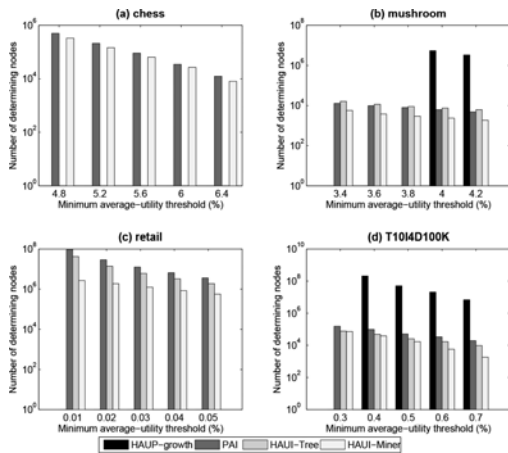


Figure 3. Number of determining nodes for various minimum average-utility values.

an extra array is attached to each node of its tree structure to keep information to be used by the mining process. When more information is stored in these arrays, the number of nodes (candidates) generated exponentially increases. No results are provided for the HAUP-growth algorithm in Figure 3(a) and Figure 3(c) since it exceeds the setup maximum time limit. From the results, we can conclude that the designed pruning strategy is efficient to greatly reduce the number of nodes of determination.

6 CONCLUSION

In this paper, an efficient average-utility (AU)-list structure is designed to store the information needed to discover HAUIs. The HAUI-Miner algorithm discovers HAUIs by exploring a set-enumeration tree using a depth-first search. An efficient pruning strategy is also developed to prune unpromising candidates early and thus reduce the search space. Substantial experiments were conducted on both real-life and synthetic datasets to evaluate the efficiency and effectiveness of the designed algorithm compared to the HAUP-growth, PAI and HAU-Tree algorithms in terms of runtime and number of determining nodes.

ACKNOWLEDGMENT

This research was partially supported by National Natural Science Foundation of China (NSFC) under grant No.61503092, and by the SGS grant No. SP2016/170, VSB-Technical University of Ostrava, Czech Republic.

REFERENCES

- Agrawal, R., T. Imielinski, & A. Swami (2005). Mining association rules between sets of items in large databases. *ACM SIGMOD Record*, 207–216.
- Agrawal, R. & R. Srikant. Quest synthetic data generator. <http://www.Almaden.ibm.com/cs/quest/syndata.html>.
- Agrawal, R. & R. Srikant (1994). Fast algorithms for mining association rules in large databases. *The International Conference on Very Large Databases.*, 487–499.
- Ahmed, C. F., S. K. Tanbeerand, B. S. Jeong, & Y. K. Lee (2009). Efficient tree structures for high utility pattern mining in incremental databases. *IEEE Transactions on Knowledge and Data Engineering* 21(12), 1708–1721.
- Chan, R., Q. Yang, & Y. D. Shen (2003). Minging high utility itemsets. *IEEE International Conference on Data Mining*, 19–26.
- Fournier-Viger, P. & J. C. W. Lin. Spmf: An open-source data mining library. <http://www.philippe-fournier-viger.com/spmf/>.
- Han, J., J. Pei, Y. Yin, & R. Mao (2004). Mining frequent patterns without candidate generation: a frequent-pattern tree approach. *Data Mining & Knowledge Discovery* 8(1), 53–87.
- Hong, T. P., C. H. Lee, & S. L. Wang (2009). Mining high average-utility itemsets. *IEEE International Conference on Systems, Man and Cybernetics*, 2526–2530.
- Lan, G. C., T. P. Hong, & V. S. Tseng (2012). Efficiently mining high average-utility itemsets with an improved upper-bound strategy. *International Journal of Information Technology & Decision Making* 11(5), 1009–1030.
- Lin, C.W., T. P. Hong, & W. H. Lu (2010). Efficiently mining high average utility itemsets with a tree structure. *Lecture Notes in Computer Science* 5990, 131–139.
- Lin, C. W., T. P. Hong, & W. H. Lu (2011). An effective tree structure for mining high utility itemsets. *Expert Systems with Applications* 38(6), 7419–7424.
- Lin, C. W., T. P. Hong, & W. H. Lu (2015). Efficient algorithms for mining up-to-date high-utility patterns. *Advanced Engineering Informatics* 29(3), 648–661.
- Liu, M. & J. Qu (2012). Mining high utility itemsets without candidate generation. *ACM International Conference on Information and Knowledge Management*, 55–64.
- Liu, Y., W. K. Liao, & A. Choudhary (2005). A two-phase algorithm for fast discovery of high utility itemsets. *Lecture Notes in Computer Science* 3518, 689–695.
- Lu, T., B. Vo, H. T. Nguyen, & T. P. Hong (2014). A new method for mining high average utility itemsets. *Lecture Notes in Computer Science* 8838, 33–42.
- Tseng, V. S., B. E. Shie, C. W. Wu, & P. S. Yu (2013). Efficient algorithms for mining high utility itemsets from transactional databases. *IEEE Transactions on Knowledge and Data Engineering* 25(8), 1772–1786.
- Yao, H., H. J. Hamilton, & C. J. Butz (2004). A foundational approach to mining itemset utilities from databases. *SIAM International Conference on Data Mining*, 215–221.
- Zihayat, M. & A. An (2014). Mining top-k high utility patterns over data streams. *Information Sciences* 285(1), 138–161.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

An extension for the mobile payment collaborative model proposed for developing countries—Egypt case study

Mahmoud Goher & Mohamed Abo Rizka

Faculty of Computer Science and Information Systems, Arab Academy for Science and Maritime Transport, Cairo, Egypt

ABSTRACT: The world witnessed a big evolution in Electronic commerce and Mobile wireless communication technologies during the last 2 decades. Besides, mobile devices penetration rate all over the world is increasing, and the dependence on mobile devices in daily personal and business routines is growing as well. However, Mobile Payments adoption and acceptance is much lower in developing countries compared with developed ones. It is highly recommended for developing countries to increase usage of Mobile Payments especially with the lack of banking services and the presence of privacy and security issues of other payment methods like Credit Cards, and cash. This paper proposes an extension to the mobile payments collaborative model. The proposed model uses the wireless telecommunication technology services of Mobile Network Operators, trust towards banks and the spread of Payments Service Providers and mediators. This model should enable customers to make remote and proximity payments securely and privately any time anywhere under the supervision of governmental authorities. As a result of implementing the proposed model, adoption and acceptance of mobile payments in developing countries might be enhanced to be suitable for a country like Egypt.

1 INTRODUCTION

Paying for goods and services has evolved between the two main players of any commercial transaction (known as “Payer” and “Payee”) from the physical exchange of goods, then evolved to notes and coins, then signing bank checks, then by sending payment card (debit or credit) details using Internet.

As Mobile Commerce (M-Commerce) is the normal successor of Electronic Commerce (E-Commerce), accordingly, Mobile Payment (MP) is the normal successor of Electronic Payment (EP) [2][5] [10]. Both EP and MP are core success or failure factors for E-Commerce and M-Commerce respectively [15]. E-commerce allowed consumers to shop while the presence at home, at any time. It is often referred to as shopping ‘anytime, anywhere’. E-Commerce requires access to a PC. M-commerce overcomes this limitation; Consumers are now able to shop anywhere. M-commerce cannot exist without payment systems. Every E-Commerce system generally includes three parts: data communication system, logistics system and electronic payment system [7].

1.1 Payments types

Based on [18][41], Payments can be categorized from perspective of: Medium (cash, check, electronic or mobile), Size (micro, mini, and macro

payments), Time of payment (prepaid, postpaid, pay now, Credit, or on installments), Domestic and international P2P transfers, and Location of purchase (physical presence in merchant’s store—i.e. proximity payments—or remote payments).

1.2 MP definition

Mobile Payments has some aliases such as (Mobile Money), (Mobile Wallet) or (Mobile Money Transfer) is defined as: “wireless transactions of a monetary value from one party to another using a mobile device whose physical form can vary from a mobile phone to any wireless enabled device (e.g. PDA, laptop, key ring, watch) which are capable of securely processing a financial transaction over a “wireless network” [33]. There is also an estimation introduced by Juniper research which stated: “*The total Mobile Payment market will be worth more than \$1.3 trillion globally by 2017 with an average annual increase of more than 23%*”. This resulted in more research intention directed to MP and its related topics and a lot of papers and scientific projects were found in the literature. MP solutions have been implemented in developing countries as an extension of financial services to the “un-banked” or “under-banked” communities. These communities are estimated to be as much as 50% of the world’s adult population, according to Financial Access’ 2009 Report “Half the World is Unbanked”.

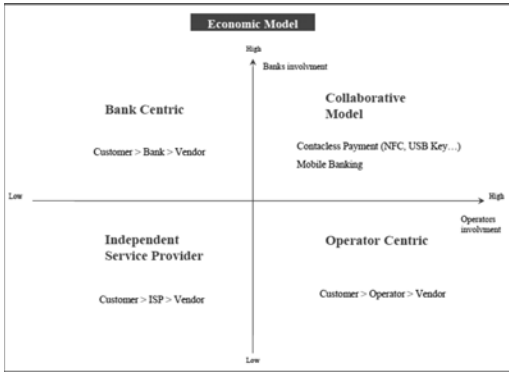


Figure 1. Economic MP models.

1.3 Mobile payment types and classification

1.3.1 Based on [18][45], MP was classified into four economic models

- Bank Centric Model
- Mobile Network Operator Centric Model
- Independent Service Provider Model
- Collaborative Model (our proposed model is an extension to this model with more technical details from analysis and design perspectives).

1.3.2 Based on [15], it proposed a classification framework for mobile payment solutions

- Smart Card Payment Schemes Driven by Financial Institutions
- Phone-based Payment Systems Operated by MNOs
- Independent Payment Schemes Using Smart Cards
- Independent Mobile Payment Solutions Using a Mobile Handset

1.3.3 Based on [40], MP was classified from technical point of view into the following platforms

- Pure SMS platform
- USSD platform
- WAP/GPRS platform
- Phone-based application platform
- SIM-based application platform
- Short range communication network platform (NFC, BT, RFID, IRDA)

1.3.4 According to [25], the MP types were classified according to the transaction flow between participants

- B2B - Business to Business
- B2C - Business to Consumer
- C2C - Consumer to Consumer
- B2G - Business to Government
- P2P - Person to Person

1.4 Characteristics of the ideal MP system which may increase the users' adoption and acceptance based on [2][10][11][25][41]

- Simplicity and Usability (i.e. Ease of use)
- Mobility
- Efficiency (i.e. performance)
- Effectiveness (i.e. usefulness)
- Universality and Interoperability
- Security, Privacy
- Risk and Trust
- Expressiveness
- Cost, Speed and Cross border payments
- Anonymity, traceability

1.5 MP main players

The rest of the paper is organized as following: in the next section, the results of the extensive literature reviews and experts' opinions were explored, information from market surveys, technical journals, product catalogs, research reports, newspapers and magazines and focusing on some previous researches and implementations worldwide, in developing countries and in Egypt. The third section presents the proposed model, stakeholders and some of the main protocols and procedures used to enable customers and merchants to make secure, easy and reliable mobile payments. Conclusion is presented in section 4, then future work and potential researches are shown in the last section.

2 RELATED WORK

This research is based on extensive literature review and, gathering information from technical journals, market surveys, research reports, and experts' opinions. A lot of challenges faced MP's

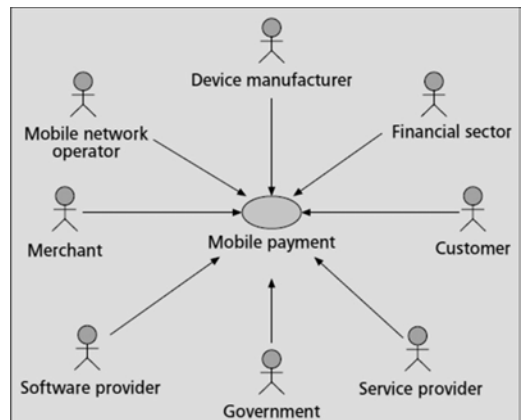


Figure 2. MP players.

security, privacy and adoption which lead to many of researches. Related researches and technologies, theories and practices will be addressed.

Paper [4] addresses the challenge of interoperability by designing a wallet service built upon concept of token. A token encapsulates a payment instruction, which can be of different types, such as instant, dated or installment payment. A token is represented by an Info-Matrix code, viz. Quick-Response (QR) code, thereby bringing in additional advantages in terms of user interactions with the wallet service. The paper also presents the lifecycle management of a token, beginning with token generation on user request, transfer of token and acknowledged receipt and encashment of a token.

From other side, [5] proposes a novel model called Mobile Payment Consortia System (MPCS) to carry out the transactions from the bank to the academic institutions for the payment of fees by students using mobile phones. MPCS provides end-to-end security using symmetric signature scheme while carrying out the payment transactions. This system increases convenience in payment processes and reduces transactional cost for both students and educational institutions. This work used the client server architecture.

In [19], the model focuses on the enhancement of privacy and non-repudiation. It introduces trusted third elements and follows new mechanism to achieve this objective. The accountability implemented by the protocol is proved by formal analysis. The model built on SEMOPS, but it introduces trusted third parties to reduce the trust dependence on the payment processor.

In [21], A New Scheme for the Electronic Coin was proposed, and the notion of ownership of electronic money was analyzed. The paper observed that a payment is the transfer of ownership of money where a payer's ownership of cash

is transferred to a payee, a proposal of a digitalized ownership transfer scheme suited for the process of coin transferal was presented. The proposed scheme models the payment system of the physical coin. Thus, the electronic coin system is required to have the following properties:

- The verifiability of ownership of a digital coin.
- The ownership of a digital coin is transferable.
- The system supports low cost, on-line payment.
- The digital coin is anonymous cash.

The main entities of the proposed payment scheme are Minting Bank and three other entities, Bank, Payer, and Payee. Minting Bank is a trusted party. Minting Bank is responsible for minting electronic coins, providing the labor for verifying the ownership of a coin and transferring the ownership of a coin from a Payer to a Payee according to the owner's instructions. Bank buys electronic coins for clients from Minting Bank to provide for the anonymity of the electronic coin owners. Any client can buy electronic coins from Bank. The proposed scheme is partitioned into the following 4 phases: The Electronic Coin Minting Phase, The Electronic Coin purchased by the Bank Phase, The Electronic Coin Withdrawal Phase, and The Electronic Coin Payment Phase. As a conclusion; in the proposed scheme, electronic coins do not depend on blind signature, which is a common technique used in the previous schemes. The scheme is based on a digital ownership transfer scheme. Therefore, it supports the reusability of a digital coin like a real coin. The main disadvantage of this model is that it didn't provide traceability in case of fraudulent and offline payments scenario doesn't exist.

Some alternatives for the NFC Secure Element (SE) were proposed in [23] which depend on the position of the SE in the handset. Also, a solution for the problem occurred recently between NFC stakeholders who encounter difficulties to define which of them should have the admin access of the SE.

Near Field Communication (NFC) is a short-range wireless technology based on RFID standard ISO 18092, ISO 14443 and ISO 15693. An NFC-enabled cell phone acts as an RFID reader to read compatible RFID tags (NFC tags), such as smart posters (using the "Reader/Writer" mode). The same cell phone can also be used as an NFC tag storing relevant data. In this case, a cell phone transforms into a digital wallet storing bank cards (money), vouchers, loyalty cards etc., at a secure place called 'Secure Element' SE (using the "Card Emulation" mode). In the peer-to-peer mode, after connecting two active NFC-enabled devices, data can be exchanged between them. SE is an encrypted, tamper-proof chip in mobile phone that stores payment keys, applications, and confidential financial data, and where execution of the payment transaction occurs.

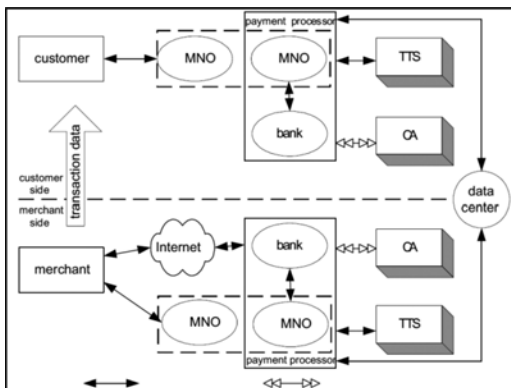


Figure 3. Model proposed by [19].

2.1 Current implementations world wide

Visa, Master card, Google Wallet (Android Pay), Paypal, SEMOPS, Apple Pay, M-Pesa, NTT DOCOMO [10], and Bitcoin.

2.2 Current implementations in Egypt

Vodafone Cash, Mobicash, Etisalat Floos, Phone Cash (NBE with Fawry), Bee Mobile.

3 THE PROPOSED MODEL

3.1 Use case diagram

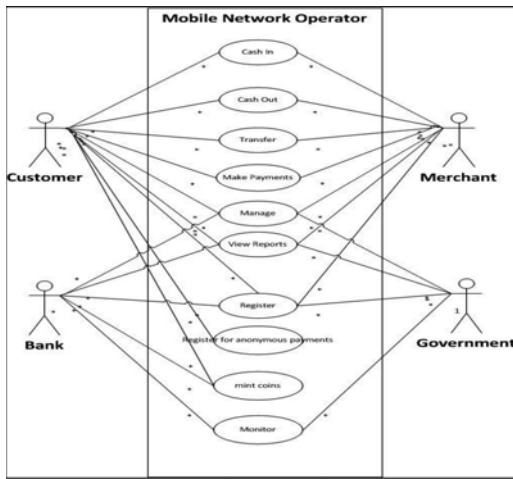


Figure 4. Use case diagram.

3.2 System architecture

The proposed model is cloud-based solution using the Service Oriented Architecture (SOA), and APIs.

3.3 Model components and main players

3.3.1 Telecommunication companies (MNOs)

Provide the secured communication infrastructure, UICC (Universal Integrated Circuit Card), STK (SIM Application Toolkit), SIM (Subscriber identity module), SMS (Short Message Service), USSD (Unstructured Supplementary Service Data), GPRS (General Packet Radio Service), WAP (Wireless Application Protocol), UMTS (Universal Mobile Telecommunications System), etc.

3.3.2 Certified banks to e-mint new e-coins

E-coin is an e-token which has its own value, serial number and security features can be minted based on the e-cash systems [12][21][28][34][35][51]. Coins

can be minted also by customers to be spent online or offline.

3.3.3 Governmental regulation authorities (e.g. NTRA, CBE)

For monitoring and making sure that all policies and processes are implemented correctly.

3.3.4 M-wallets owners

Customers and Merchants are main players of the model. Customers can use the mobile payment system to Cash In, Cash Out, Transfer money, Make online/offline payments (Topup Airtime, pay bills and utilities, and pay for physical and online merchants). Customers can also manage their accounts, and their sub accounts (using the RBAC—Role Based Access Control—approach). Payments can be done anonymously.

3.4 Proposed payment scenarios

- Over The Counter (OTC) using Bluetooth, WIFI, NFC and Infrared. This can be used by users who own smart phones.
- Over The Air (OTA) using SMS, USSD, GPRS, WAP, and UMTS. This can be used by normal phones users.

3.5 Security analysis

- Blind signatures and group signatures are used to ensure anonymity and traceability (under certain circumstances and regulations monitored by government) [21][28][34][51].
- Overspending and double spending detection and prevention techniques can be used by Governmental Regulation Authorities as mentioned in [27] [46]. Overspending means that customer account does not have enough balance to make a payment. Double spending is that customer uses the same coins to make payments.

3.6 Usage and Applications of the proposed model

The proposed model can be used in the following applications:

- Digital coin can be minted by any user who is authorized to do. This minted coin can be used for occasional events (or closed places).
- By the government to allow citizens to pay certain list of products with lower prices from certain merchants (Tamween).
- It can be used also by work owners to distribute wages to its employees.

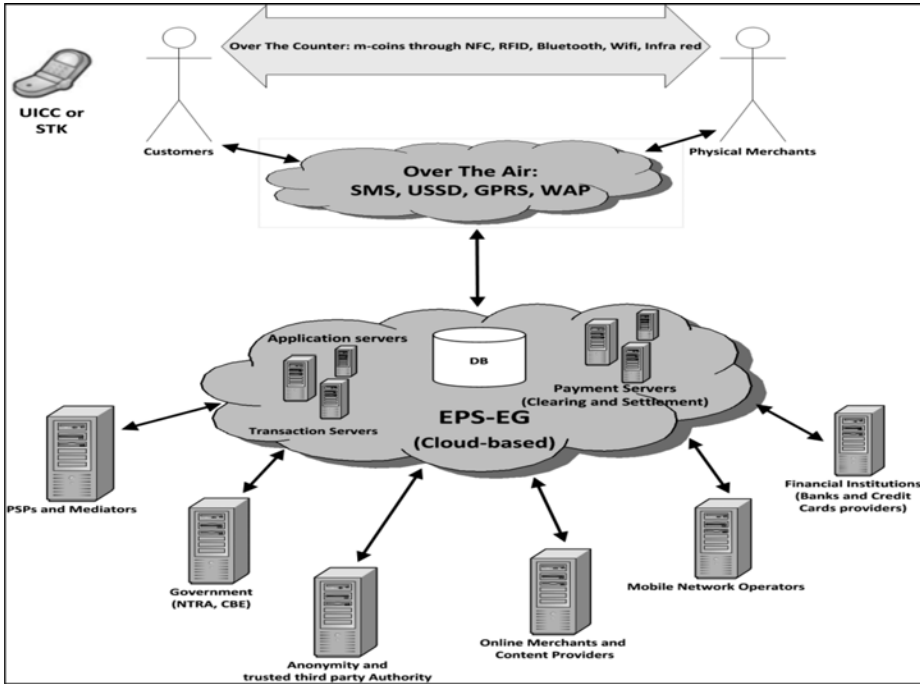


Figure 5. Conceptual diagram.

- Domestic and international P2P payments are supported even if the two peers are subscribed to different MNOs.
- Parents to control the payments of their children in schools selecting what product categories they can buy.
- The fuel redemption system in Egypt (instead of the card based solution).

4 CONCLUSION

By implementing the proposed model in a developing country like Egypt, some advantages can be satisfied. It is a generic payment system, which uses a lightweight protocols for OTA money and credentials transfer operations supported by the MNO. It also allows Micro, Mini and Macro payments under certain regulations specified by each government. The proposed model also allows parental control management by users who have the appropriate rights to do on certain selected list of sub-users or sub-wallets. Banked and unbanked customers will be allowed to make payments anytime, anywhere using their smart phones on Android, IOS, Windows, BlackBerry, Web OS, etc. The system supports Anonymity and Privacy of users, however it supports Traceability in some

scenarios when a fraud case occurs (like double spending or overspending). Payments can be Online (through MNO—UICC), Offline through NFC and other proximity wireless technologies. Coins can be used for specific events were the issuer can allocate this coins type for certain event (cannot be used in other events).

5 FUTURE WORK

The future work will be focused on development and implementation of the proposed model in Egypt with cooperation with government, MNOs, banks, payment service providers and payment mediators and aggregators. The most challenging part which should be considered in future researches is the fair distribution of roles, revenues of the stakeholders of collaborative model.

REFERENCES

- [1] Ok, Kerem, et al. "Current benefits and future directions of NFC services." Education and Management Technology (ICEMT), 2010 International Conference on. IEEE, 2010.
- [2] Carr, Mahil. "Mobile payment systems and services: an introduction." Mobile Payment Forum. 2007.

- [3] Dahlberg, Tomi, et al. "Past, present and future of mobile payments research: A literature review." *Electronic Commerce Research and Applications* 7.2 (2008): 165–181.
- [4] De, Pradipta, et al. "Towards an interoperable mobile wallet service." *Emerging Technologies for a Smarter World (CEWIT), 2013 10th International Conference and Expo on.* IEEE, 2013.
- [5] Kumar, S. Britto R., and S. Albert Rabara. "A framework for mobile payment consortia system (MPCS)." *Computer Science and Software Engineering, 2008 International Conference on.* Vol. 2. IEEE, 2008.
- [6] Zmijewska, Agnieszka, and Elaine Lawrence. "Implementation models in mobile payments." *ACST*. 2006.
- [7] Zmijewska, Agnieszka, Elaine Lawrence, and Robert Steele. "Towards Understanding of Factors Influencing User Acceptance of Mobile Payment Systems." *ICWI*. 2004.
- [8] Schierz, Paul Gerhardt, Oliver Schilke, and Bernd W. Wirtz. "Understanding consumer acceptance of mobile payment services: An empirical analysis." *Electronic Commerce Research and Applications* 9.3 (2010): 209–216.
- [9] Dahlberg, Tomi, Niina Mallat, and Anssi Öörni. "Trust enhanced technology acceptance model consumer acceptance of mobile payment solutions: Tentative evidence." *Stockholm Mobility Roundtable* (2003): 22–23.
- [10] Karnouskos, Stamatis. "Mobile payment: a journey through existing procedures and standardization initiatives." *Communications Surveys & Tutorials, IEEE* 6.4 (2004): 44–66.
- [11] Mallat, Niina. "Exploring consumer adoption of mobile payments—A qualitative study." *The Journal of Strategic Information Systems* 16.4 (2007): 413–432.
- [12] Meng, Bo, and Qianxing Xiong. "Research on electronic payment model." *Computer Supported Cooperative Work in Design, 2004. Proceedings. The 8th International Conference on.* Vol. 1. IEEE, 2004.
- [13] Delic, Natali, and Ana Vukasinovic. "Mobile payment solution-symbiosis between banks, application service providers and mobile network operators." *Information Technology: New Generations, 2006. ITNG 2006. Third International Conference on.* IEEE, 2006.
- [14] Petrova, Krassie. "Mobile payment: Towards a customer-centric model." *Web Information Systems Engineering—WISE 2008 Workshops.* Springer Berlin Heidelberg, 2008.
- [15] Ondrus, Jan, and Yves Pigneur. "Towards a holistic analysis of mobile payments: A multiple perspectives approach." *Electronic Commerce Research and Applications* 5.3 (2006): 246–257.
- [16] Isaac, Jesús Tellez, and José Sierra Cámara. "Anonymous payment in a client centric model for digital ecosystems." *Digital EcoSystems and Technologies Conference, 2007. DEST'07. Inaugural IEEE-IES.* IEEE, 2007.
- [17] Au, Yoris A., and Robert J. Kauffman. "The economics of mobile payments: Understanding stakeholder issues for an emerging financial technology application." *Electronic Commerce Research and Applications* 7.2 (2008): 141–164.
- [18] Van Bossuyt, Michaël, and Leo Van Hove. "Mobile payment models and their implications for NextGen MSPs." *info* 9.5 (2007): 31–43.
- [19] Liu, Jun, Jianxin Liao, and Xiaomin Zhu. "A system model and protocol for mobile payment." *e-Business Engineering, 2005. ICEBE 2005. IEEE International Conference on.* IEEE, 2005.
- [20] Fun, Tan Soo, et al. "A lightweight and private mobile payment protocol by using mobile network operator." *Computer and Communication Engineering, 2008. ICCCE 2008. International Conference on.* IEEE, 2008.
- [21] Chan, Chao-Wen, and Chin-Chen Chang. "A new scheme for the electronic coin." *e-Business Engineering, 2006. ICEBE'06. IEEE International Conference on.* IEEE, 2006.
- [22] Kungpisdan, Supakorn, Bala Srinivasan, and Phu Dung Le. "A secure account-based mobile payment protocol." *Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004. International Conference on.* Vol. 1. IEEE, 2004.
- [23] Reveilhac, Marie, and Marc Pasquet. "Promising secure element alternatives for NFC technology." *Near Field Communication, 2009. NFC'09. First International Workshop on.* IEEE, 2009.
- [24] Hu, Jhe-Yi, et al. "Android-based mobile payment service protected by 3-factor authentication and virtual private ad hoc networking." *Computing, Communications and Applications Conference (ComComAp), 2012. IEEE, 2012.*
- [25] Singh, Basudeo, and K. S. Jasmine. "Comparative study on various methods and types of mobile payment system." *Advances in Mobile Network, Communication and its Applications (MNCAPPS), 2012 International Conference on.* IEEE, 2012.
- [26] Ondrus, Jan, and Yves Pigneur. "Near field communication: an assessment for future payment systems." *Information Systems and E-Business Management* 7.3 (2009): 347–361.
- [27] Lin, Phone, et al. "A secure mobile electronic payment architecture platform for wireless mobile networks." *Wireless Communications, IEEE Transactions on* 7.7 (2008): 2705–2713.
- [28] Chaum, David. "Blind signatures for untraceable payments." *Advances in cryptology.* Springer US, 1983.
- [29] IDA-Pay an innovative micro-payment system based on NFC technology for Android mobile devices
- [30] Haselsteiner, Ernst, and Klemens Breitfuß. "Security in near field communication (NFC)." *Workshop on RFID Security RFIDSec.* 2006.
- [31] Kadhiwal, Saleem, and Anwar Usman Shaheed Zulfiquar. "Analysis of mobile payment security measures and different standards." *Computer Fraud & Security* 2007.6 (2007): 12–16.
- [32] Hassinen, Marko, Konstantin Hyppönen, and Elena Trichina. "Utilizing national public-key infrastructure in mobile payment systems." *Electronic Commerce Research and Applications* 7.2 (2008): 214–231.
- [33] Chaix, Laetitia, and Dominique Torre. "The dual role of mobile payment in developing countries."

- No. 2015–01. Groupe de REcherche en Droit, Economie, Gestion (GREDEG CNRS), University of Nice Sophia Antipolis, 2015.
- [34] Miers, Ian, et al. “ZeroCoin: Anonymously distributed e-cash from bitcoin.” *Security and Privacy (SP)*, 2013 IEEE Symposium on. IEEE, 2013.
- [35] Wang, Da-Xing, and Ji-Kai Teng. “Research and analysis of electronic cash payment system.” *Educational and Information Technology (ICEIT)*, 2010 International Conference on. Vol. 3. IEEE, 2010.
- [36] Mathew, Mary, N. Balakrishnan, and S. Pratheeba. “A study on the success potential of multiple mobile payment technologies.” *Technology Management for Global Economic Growth (PICMET)*, 2010 Proceedings of PICMET’10. IEEE, 2010.
- [37] Pousttchi, Key. “Conditions for acceptance and usage of mobile payment procedures.” (2003): 201–210.
- [38] Gusev, M. A. R. J. A. N., Ljupco Antovski, and Goce Armenski. “Models of mobile payments.” *Proceedings of WSEAS ICOMIV* (2002): 3581–3586.
- [39] Meng, Jian, and Liang Ye. “Secure mobile payment model based on wap.” *Wireless Communications, Networking and Mobile Computing*, 2008. WiCOM’08. 4th International Conference on. IEEE, 2008.
- [40] Tehrani, Mohammad, et al. “A survey of system platforms for mobile payment.” *Management of e-Commerce and e-Government (ICMeCG)*, 2010 Fourth International Conference on. IEEE, 2010.
- [41] Ramezani, Elham. “Mobile Payment.” *Lecture E-Business Technologies, BCM1* (2008).
- [42] McKitterick, David, and Jim Dowling. “State of the art review of mobile payment technology.” Retrieved September 14, 2003 (2003): 2003–24.
- [43] <http://www.nfcworld.com/2012/10/11/318353/ntt-docomo-to-take-japanese-mobile-wallet-global/>
- [44] Ubaya, Huda. “Design of Prototype Payment Application System With Near Field Communication (NFC) Technology based on Android.” *Computer Engineering and Applications Journal (ComEngApp)* 1.1 (2012): 1–12.
- [45] Chaix, Laetitia, and Dominique Torre. “Four models for mobile payments.” *University Nice Sophia-Antipolis, JEL Classification E 42* (2011): O33.
- [46] Asokan, Nadarajah, et al. “The state of the art in electronic payment systems.” *Computer* 30.9 (1997): 28–35.
- [47] Lee, Zon-Yau, Hsiao-Cheng Yu, and Pei-Jen Ku. “An analysis and comparison of different types of electronic payment systems.” *Management of Engineering and Technology*, 2001. PICMET’01. Portland International Conference on. IEEE, 2001.
- [48] Pouralinzar, Behzad. “The System for Secure Mobile Payment Transactions.” (2013).
- [49] Zhu, Yunpu, and Jacqueline E. Rice. “A lightweight architecture for secure two-party mobile payment.” *Computational Science and Engineering*, 2009. CSE’09. International Conference on. Vol. 2. IEEE, 2009.
- [50] Cheng, Hsu-Chen, et al. “A generic model for NFC-based mobile commerce.” *Advanced Communication Technology*, 2009. ICACT 2009. 11th International Conference on. Vol. 3. IEEE, 2009.
- [51] Zhang, Ling, Jianping Yin, and Mengjun Li. “A novel off-line anonymous and divisible digital cash protocol utilizing smart card for mobile payment.” *Communications and Networking in China*, 2006. ChinaCom’06. First International Conference on. IEEE, 2006.
- [52] Wang, Hua, Jinli Cao, and Yanchun Zhang. “A flexible payment scheme and its role-based access control.” *Knowledge and Data Engineering, IEEE Transactions on* 17.3 (2005): 425–436.
- [53] Ho, Henry, Simon Fong, and Zhuang Yan. “User acceptance testing of mobile payment in various scenarios.” *e-Business Engineering*, 2008. ICEBE’08. IEEE International Conference on. IEEE, 2008.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Measuring adolescents awareness security of internet a critical analysis of the internet and adolescents self-injury

Fai Ben Salamh

College of Computing Sciences and Engineering, Kuwait University, Kuwait

ABSTRACT: The internet can be both a source of safety and support and a source of very real danger for Adolescents who are struggling with self-harm. The term self-harm is a prevalent behavioral problem, which has series study reports on the diagnostic correlates of adolescents with a recent history. The term self-harm is commonly used to describe a wide range of behaviors and intentions including attempted hanging, impulsive self-poisoning, and superficial cutting in response to intolerable tension. The rising of global rate of related psychological weakness among adolescence have obtained a considerable attention, where studies shows that adolescents use the Internet for the purpose of connecting with others at higher rates than any other age group, so a better understanding of how Internet use affects their social and emotional development is an important line of scientific inquiry. The identification of successful prevention initiatives aimed at adolescences and those at especially high risk, are fundamental needs, prevention of self-harm needs both global measures targeted adolescence category in general and aimed initiatives focused on high-risk groups. This paper summarizes the current scientific knowledge regarding the relation between internet and self-injury. Clinical features, epidemiology, assessment methods, and existing treatments of self-injury. The role of the primary care physician in the treatment of patients who self-injure is specifically outlined.

Keywords: self-injury, self-harm, cyber bullying, internet safety and Adolescents

1 BACKGROUND

People have all sorts of feelings about self-injury including fear, anger, frustration, helplessness and worry. Being someone who self-injury or being close to someone who self-injury is tough, understanding why someone close to you hurts themselves can be very difficult, even scary. Self-injury is a way of expressing and dealing with deep distress and emotional pain. Today it shows that self-injury is a relatively common phenomenon in adolescence, it does affect all age groups, but studies have consistently shown that it's increasing among adolescence. The action is used for reasons that relate to reducing distressing affect, where it is their only way they know how to cope with their bad feelings. Researchers have found that in the past two decades self-injury phenomena is more prevalent in adolescents, and studies indicated that internet is linked to an increased risk of self-injury among adolescents. Adolescents who are addicted to the Internet are more likely to engage in self-harm behavior, according to many studies. Internet addiction as a distinct set of behaviors, however, "Many studies have reported associations between Internet addiction, psychiatric symptoms and depression among adolescents." Studies shows that adolescences are at risk of self-injury where they are often online

for longer period than other ages, as, much attention has been focused on adolescent risk behaviors. Self-injury is consider as a dangerous behavior that is different from suicidal behavior but is associated with increased risk of suicide attempts. {11}

News media have alarming' numbers of adolescences are being exposed to self-harm which could encourage them to hurt themselves. According to the Center for Adolescent Health at the Murdoch Children's Research Institute concluded that around 1 in 12 youth self-harm, which is an alarming statistic and may be a bit unbelievable for some parents to understand. A large number of adolescences at Newport Academy who have had cutting or other self-harm issues that many of their loved ones had no idea about. One of the biggest and most heartbreaking problems is these adolescences are very skilled at hiding their pain. The inflicted wounds stems from depression, anxiety, stress or pressure. {10}

2 LITREATURE REVIEWS

There are many studies that have demonstrated a strong relationship between Internet and self-harm phenomena among adolescences. Studies reported the related between internet addiction, and

depression among adolescents. The results suggested a “strong and significant” association between internet addiction and self-injurious behavior. Internet among adolescents presents a crucial argument for assessment of Internet use in general. Where there is a strong link between Adolescents how’s using internet forums and an increased risk of suicide. {5} Internet as a dirt road without a sheriff, {3} however today, it’s common to see blogs about cutting, “how-to” guides about suicide and communities where users engage in or support such behavior. Internet is linked to an increased risk of suicide and self-injury among adolescents. Self-injury is a common health problem for adolescents. And found that about 25,000 Australians adolescences are admitted to hospital each year because of self-harm. {14}

And, approximately 93% of American adolescent’s between ages 12 to 17, use the Internet, and nearly two-thirds of adolescent Internet users go online daily, and these numbers are growing every day, and the number of websites intended for or about people who self-injure has increased too. {13} Moreover, more than 500 message boards focused on self-injury, and observed the related between the increase in self-injury websites and the growth in self-injury awareness in society. {16} Internet message boards provide a strong reasons for bringing adolescents who self-injure together.

Also, Non-Suicidal Self-Injury (NSSI) has been recognized as a significant mental health problem in adolescence with high prevalence rates. {12}

The most common methods of self-injuries in Adolescents is (Cutting), with more than 70%, then the second methods is the Self-poisoning. Also 5% of self-injuries in hospitals will have committed suicide within 9 years. {2}

Rating self-injuries are vary greatly between countries, and shown that 5–9% of Adolescents in western countries are having self-injuries within the previous year. Also there are many cultural aspects of some societies may protect against suicide and self-injury and explain some of the international variation in rates of these events. Risk of repetition of self-injury increased every day. {7}

BBC Radio Wales conducted that UK has one of the highest rates of self-harm in Europe. These results demonstrate with some variability the significant rates of NSSI in youth in a cross-country study. {1} On the other hand, (18.14%) of Jordanian adolescences reporting having engaged in NSSI at least once in their lifetime, which provide an empirical evidence that adolescent engagement in NSSI occurs at similar prevalence levels in Jordan, relative to North American samples. {6}

Also, 14–17% of England adolescents have self-injured, as have 4% of adults, and, these numbers increase dramatically to 20% among adult

psychiatric and 40–80% of adolescent psychiatric patients, where, the first survey was conducted in 2005 in a general population of adolescents and young adults in the U. S. Using a web-based survey of students from two colleges in the northeast, the team examined self-reports of self-injurious practices, age of onset, forms, severity, intention, and help seeking behavior. Based on responses from 3,069 students, the finding contributed to the emerging profile of self-injury: that it shows that it was happening in individuals who had never been in therapy for any reason, was surprisingly widespread, and that few disclosed their behavior and sought help. {15}

In another study that conducted in Canada found that 12.4–17% of adolescences reported self-harm, where, 77% of them were female, and 40% were repeat harmers, 75% said it was their idea, while 29% got the idea from friends, 2% from family, 15% from the media and 12% said that they have read about it. {2}

Moreover, more than half of adolescents between ages from 11 to 14, indicated that they had shared pictures of people self-harming on social media, and more than half of adolescences in this age group who viewed images of self-harm online said they felt like hurting themselves afterwards. {4}

The mental health charities warned of a disturbing trend of young people posting self-harming ‘selfies’ online, due to the rise in popularity of apps such as Instagram and Snap chat, that warning was posted in January 2014. {8}

In the other hand, BBC Radio Wales declared about a study by Cardiff University that has found that adolescents, aged between 10 to 19, self-harm has increased, while in 2013–14 were (1,542) of adolescents went to hospital because of self-harm compared to (1,134) between 2012 and 2013. {1} Studies shows that these adolescents not only use social media to share images of self-harm—some go further and use the images as part of their own self-harming behavior.

Where, more than (1,600) of adolescents in China were moderately addicted to the Internet, 2.4 times more likely to have self-injured themselves than those with normal Internet habits, where they were severely addicted to the Internet nearly five times as likely to injure themselves. {4}

In China 1618 adolescents aged between ages 13 to 18 found that about one in six reported some form of self-injury such as hitting, burning or cutting themselves. Just over one in ten reported moderate or severe internet addiction: admitting to feeling depressed or moody when ‘off-line’, and fantasizing about the internet when away from a computer. When the results were compared, where, students were twice as likely to report high levels of self-harm, if they also showed signs of internet addiction. {9}

3 THE REASERCHER CRITICAL ANALYSIS

The studies above found that Internet addiction on its own can lead to Adolescents harming themselves. Recent data and self-injury statistics show that (Cutting, Burning, Inserting objects into skin, purposely burning or breaking ones bones) are common behaviors among people who engage in self-harm. The self-harming phenomenon has become more visible in society in recent years. Self-injury is a major concern because it is likely the symptom of other serious underlying causes, so treatment is important. Seek professional help if you, or others you know, have difficulties with self-injury and reduction in serious self-injury comes from reducing the contributing factors such as sexual abuse, child abuse, and emotional abuse which occurs in childhood and adolescence.

Self-injury statistics show that this disturbing phenomenon is a real and present danger to adolescences worldwide, especially in developed countries, such as the UK, U. S. and those in Western Europe. Frequently, untreated depression and other mental health challenges create an environment of despair that leads people to cope with these challenges in unhealthy ways. The pressure of online environment has been blamed for a huge rise in self-injuries among adolescents. It is the pressures of the modern world. but we are not saying that all Adolescents who goes on the internet increase their risk of suicide or self-injury. We are talking about the vulnerable categories of adolescents who are going online specifically to find out more about harming themselves or because they are considering suicide already. So the internet have both the negative and positive side and it's depend on how you use it.

4 CONCLUSION AND RECOMMENDATION

Self-harm is a particularly adolescent phenomenon and critical public health issue that need assisting the patient in identifying motivations for treatment and treatment options, and provision of long-term behavioral and risk monitoring. Adolescence are often feel misunderstood and ashamed, and, are most likely to disclose their feeling and motivations which may lead to self-injury. However, estimating the extent of the problem has proved its difficulties. Studies shows that trying to understand more about the meaning and function of self-harm would help to design more effective public health strategies to prevent problems arising.

Moreover, physicians and family support commonly have a strong relationship with a patient that has developed over several years. Through,

trusting and confidential that will lead to a great deal of trust on the part of the patient. Where, parent's and family supports are substantial in implementing treatment as well as, recommendations. On the other hand, Pharmacological treatment should be recommended through developing an understanding of the behavior, and communicating the understanding of patients' experiences, to assist the adolescent in understanding the concept of self-harm, and try to employ more adaptive coping strategies. Where, physicians got a special technique that consider as a useful set of tools through motivational interviewing techniques.

The following are examples of realization questions based on (MI) techniques that can enhance a physician's understanding reasons of self-injury from the patient's point of view, as well as, facilitate adequate discussion, that help a patient to begin thinking about getting help for his/her self-injury:

1. Do you have any previous affect in your life?
2. Is self-injury seems a function for you?
3. Do you have any motivations to stop your self-injury at this moment?
4. Do you think that it is difficult to handle life stress without self-injury?
5. What do you think from the large options available will help in stopping self-injury?

Recommendations and key considerations for the prevention of social contagion of NSSI and how to deal with the Adolescents who already self-injury: They recommend to try to limit the adolescents time spent online and take the computer out of their rooms, ask them about it, and get them help. And in case if they already get Self-Injury, you must to see a counselor indeed therapy helps individuals to reduce self-injurious behavior, process and express emotions and feel better about themself. Also parents should try to explore why their kids self-injure, and what purpose it serves for them can help to stop this behavior. Of course choosing Healthy Coping Activities Picking a healthy activity makes them feel better, they also suggested an ides called "The Magic Box" where they can in a box, put their list of coping options, a favorite movie, a few good books, an exercise DVD, soothing music, drawing supplies and journal. Whenever they are experiencing the urge to self-injure (or a negative emotion), they can get the box out and choose a healthy option instead. They recommend to avoid anything that promotes self-injury, this might mean avoiding Internet sites or even friends who glorify intentionally injuring yourself. (Cutting and Self-Injury: Finding Better Ways to Cope), also must develop a clinical understanding of social contagion and its significant impact on the adolescent population through training and further research, and develop awareness

of appropriate environments to help them how to discuss their self-injury stories, such as individual therapy sessions. Also they suggested to prohibiting graphic detail of NSSI at the onset of group therapy, and incorporating strength-based strategies that encourage healthy coping behaviors in treatment, are summarized finally in instructing the adolescences to share stories of healing and healthy coping behaviors to decrease the opportunity for contagion, while inspiring altruistic motives in a group environment. {13}

The researchers suggest that, in future, clinical assessments of such young people should include questions about the online content they have viewed. Where, this validation of a patient's feelings can foster a stronger patient-clinician relationship. As, it can assist in a greater progress in therapy. For psychotherapists. On the other hand, pushing social media sites to take larger steps in removing self-harm content off their platforms, parents need to be involved in their adolescence's lives as a way to combat depression or other mental health problems.

In conclusion Studies shows that internet use may exert both positive and negative effects on Adolescents at risk of self-harm. Self-injurious behavior can be attributed to many different factors, such as depression, stressful life events or family problems. Also some studies found that internet supported and connect socially isolated people, and helping them to cope. However, other studies concluded that Adolescents who went online to find out more about self-injury and suicide were exposed to violent imagery and acted out what they had seen online. The review finds that internet use is linked with more violent methods of self-injury. Also NSSI have a comparable prevalence in studies with adolescents from different countries.

Interventions found to be helpful in these studies include psychosocial assessment in the emergency department, therapeutic assessment by mental health providers, working to discover the meaning of NSSI to the individual patient, continuity of care, encouraging secondary education, and providing patient education including first-aid training. During adolescence, self-cutting and other self-harm are common. Adolescents who have self-cutting or harm themselves have wide-ranging problems in their lives. The specific characteristics of these phenomena need further investigation. However, few treatments have been studied in the adolescent population.

REFERENCES

[1] BBC News, (2015). 'Direct link' between self-harming and the internet, <http://www.bbc.com/news/uk-wales-31878391>

[2] Betty, F. (2010). *Adolescent self-harm*, Cutting away the pain. <http://www.parkhurstexchange.com/clinical-reviews/oct10/adolescent-self-harm>.

[3] Corbly, L. (2015). The Internet and depression: How seeking community can cause harm. <http://national.deseretnews.com/article/3627/the-internet-and-depression-how-seeking-community-can-cause-harm.html>.

[4] Davis, L. (2011). The Internet May Be Causing More Harm to Your Children Than You Think, Video Game Addiction. <http://www.video-game-addiction.org/video-game-addiction-articles/internet-may-be-causing-more-harm-to-your-children-than-you-think.htm>.

[5] Jerome, C. (2013). Internet use link to increase in self-harm, Young Minds. http://www.youngminds.org.uk/news/blog/1682_internet_use_link_to_increase_in_self-harm.

[6] Hanania JW1, Heath NL, Emery AA, Toste JR, Daoud FA. (2014), Non-Suicidal Self-Injury Among Adolescents in Amman, Jordan. <http://www.ncbi.nlm.nih.gov/pubmed/25058810>.

[7] Keren, S. (2010). Helping those who self-harm, <http://www.sciencedirect.com/science/article/pii/S0140673605676003>.

[8] Madlen, D. (2015). Is the Internet Encouraging children to self-harm Alarming number are exposed to graphic images online, charities warn. Mail online.

[9] Miller, N. (2015). Internet addiction linked to self-harming among teens. <http://www.addictioninfo.org/articles/3946/1/Internet-addiction-linked-to-self-harming-among-teens/Page1.html>.

[10] Monroe, J. (2014). Social Media Spreading Self-Harm Behavior Amongst Teens, Huff post Healthy Living. http://www.huffingtonpost.com/jamison-monroe-jr/social-media-spreading-se_b_5166748.html.

[11] Oxford University (2013). How internet affects young people at risk of self-harm or suicide. <http://www.ox.ac.uk/news/2013-10-31-how-internet-affects-young-people-risk-self-harm-or-suicide>.

[12] Plener P, Fischer C, Albon T, Rollett B, Nixon M, Groschwitz R, Schmind M. (2013), Adolescent non-suicidal self-injury (NSSI) in German-speaking countries: comparing prevalence rates from three community samples. <http://link.springer.com/article/10.1007%2Fs00127-012-0645-z>.

[13] Richardson B., Surmitis K., (2014). Responding to the rise in self-injury among youth. Counselling Today. Web Site: <http://ct.counseling.org/2014/10/responding-to-the-rise-in-self-injury-among-youth/>

[14] Strickland, M. (2006). An Information Booklet for Young People Who Self Harm & Those Who Care for them. Logan-Beaudesert Mental Health Service Queensland Health, Graphic Design & Production by Speak Out Ltd. http://www.decd.sa.gov.au/speced2/files/pages/chess/hsp/information/revise_selfharm_finalweb.pdf.

[15] Whitlock, J. (2009). The cutting edge: non-suicidal self-injury in adolescence. Retrieved from Cornell University, Act for Youth Center of Excellence website: http://www.actforyouth.net/resources/rf/rf_nssi_1209.pdf.

[16] Whitlock, J., Eckenrode, J., & Silverman, D. (2006). Self-injurious behaviors in a college population. *Pediatrics*, 117(6), 1939–1948.

Optimal throughput of time power switching relaying protocol with imperfect channel state information

Hoang-Sy Nguyen & Dinh-Thuan Do

*Wireless Communications Research Group, Faculty of Electrical and Electronics Engineering,
Ton Duc Thang University, Ho Chi Minh City, Vietnam*

Anh-Hoa Bui Thi

Eastern International University, Vietnam

Miroslav Voznak

*Faculty of Electrical Engineering and Computer Science, Technical University of Ostrava, Ostrava,
Czech Republic*

ABSTRACT: There are a lot of analysis on Wireless Powered Communication Networks (WPCN) that has become research tendency for 5th generation networks. It helps wireless devices to enhance battery life and reliability. In this paper, we illustrate two transmission modes that are delay-tolerant and delay-limited which look after two functions of wireless data and power transfer. We also depict the optimal tradeoff into minimum outage probability or maximum throughput that is produced by an energy harvesting plan. However, the change of energy harvesting can remarkably rely on the correctness of the channel estimation algorithm that is used for data and energy transfer. In addition, we present the throughput and outage probability during a WPCN. Finally, this paper analyzes the Time Power Switching Relaying (TPSR) schemes performance of the dedication of power sources that the ergodic capacity and delay-limited capacity might be maximized. The results in numerical analysis show the accuracy of the derivative of expressions, giving the design insights into WPCNs and depicts the importance of optimal throughput during a huge amount of channel estimation error.

Keywords: Energy harvesting, amplify-and-forward, channel state information, cooperative communications, throughput

1 INTRODUCTION

In the present, the purpose of 5G cellular networks that harvested energy has been the important part, is not use wires anymore. In (Varshney 2008) presented Wireless Powered Communication Network (WPCN) and it showed that the relevance of the trustworthy data to the transmission ratio on an interference channel. The RF harvested energy is described as autonomous operation and self-sustainability asset; and it utilizes a term of redundancy power of the natural environment. In (Grover & Sahai 2010), the instance of frequency-selective channels is enhanced by this research. In addition, in (Varshney 2008, Grover & Sahai 2010) the extract energy and data at once and signal might be deciphered by the receiver, however, it is not apply on the current high-tech circuit growth. Furthermore, WPCN is perceptible because of the ability to supply the portability of wireless point

and the essential of data network systems is the Radio-Frequency (RF) of surrounding environment. In existing point-to-point networks (Zhou, Zhang, & Ho 2013), in (Ng, Lo, & Schober 2013, Visser & Vullers 2013, among others) the vital attainment is the wireless transporting power of 1uW and 3.5 mW between 0.6 and 11 meters from RF signals. In order to get the benefit of spectrum efficiency, in (Hyungsik & Rui 2014) at base station applied full-duplex technology thanks to applications of WPCN. There are many absorbing results in (Rubio & Pascual-Iserte, Chalise, Wing-Kin, Zhang, Suraweera, & Amin 2013, among others) illustrate that the wireless energy transfer sacrifices by the transmission rate. The relay-assisted systems with power transfer ability has some existing major ways in publications, for example, 1) the power from the radiated signal event and origin signal are salvaged by the employed relay in WPCN scenarios (Zhiguo, Perlaza, Esnaola,& Poor 2014); 2)

Multihop power transfer scenarios in which a relay (Chalise, Wing-Kin, Zhang, Suraweera, & Amin 2013) or many relays (De & Singhal 2012) transferred the energy to remote terminals.

Then in (Xiang & Tao 2012, Tourki & Alouini 2013, among others), they discuss about the obsolete Channel State Information (CSI) at the transmitter, and about an opportunistic regenerative relaying that is used to keep the quality of service (QoS) of the secondary link operating whenever it declines below an acceptable level in an underlay cognitive network. In (Laneman & Tse 2001), the author also mentioned that co-channel interference and energy are restricted by modern wireless ad-hoc networks by deploying multi-hop transmission through intermediate relay mobiles.

In this paper, we will investigate and derive the throughput performance and outage for the systems which has one-way replay with an imperfect of CSI via testing both transmissions functions of AF network to discover optimal throughput.

The major contributions to this study are summed up as below:

- This paper derive the ability of concurrent wireless harvested energy and data processing at the power constrained relay is enabled via TPSR protocol.
- It depicts the analytical expressions of the throughput at the target for two transmission modes: delay-limited and delay-tolerant.
- In numerical results reveal an imperfect of CSI has effect on the performance of systems which includes throughput and outage probability. The empirical design insights into the impact on various parameters on the performance of system is given by the provided expressions.

In addition, this paper is put in order as below. In Section 2, we will depict the fundamental preliminaries and considerable system model that are associated with the proposed harvested energy and harvested power assisted replay. Next, we will present the throughput and outage probability for various transmission modes in Section 3. Then in Section 4, we will show numerical results that are analyzed and investigated. Finally, Section 5 will provide a conclusion.

2 SYSTEM MODEL

The wireless communication system of AF relaying is considered in which (S) the source node transfer the data to (D) destination node, through a relay node (R) which is power constrained intermediate. The assumed quasi-static of Rayleigh fading channel obtains between (S) and (R) at the first hop and between (R) and (D) at the second hop are described by h and g .

Furthermore, d_1 and d_2 denoted the interval of two hops in the relay system. The power constrained node that the source signal is harvested energy and then the data at source is forwarded to destination by using the energy harvesting. It is supposed that every received block which requires the lowest level energy of received signal, carries out the data transfer and harvested power. The single antenna model of source, destination and relay is designed at below.

The estimation of CSI can support WPCN because the deployment Request-To-Send/Clear-To-Send (RTS/CTS) is based on the algorithm of channel estimation. The first and second time slots in the entire block time (T) are working beneath shared transmitted energy of P from (S), next at the third time slot, information processing uses the energy harvesting. In addition, T is block time that the channel is presumed stable in time slot committed for data transmission and the fraction for harvested power from source signal is α , where $0 \leq \alpha \leq 1$. In the first time slot, it occupies αT time whilst $(1-\alpha)T$ is remaining time that is used for data transmission in both hops. In the remaining block time, $(1-\alpha)T/2$ is used for (S) to (R) data transmission and another $(1-\alpha)T/2$ is used for (R) to (D) data transmission. Whereas the entire transmitted energy from (S) has two components, βP used for power harvesting component and $(1-\beta)P$ served for data processing at (R).

In the proposition of power constrained relay-assisted transmission, the baseband Additive White Gaussian Noise (AWGN) adds the received signal because of the obtaining antenna at (R).

The received signal at (R), $y_r(t)$, is presented by

$$y_r(t) = \sqrt{d_1^{-m} P_s} (h + \Delta h) s(t) + n^{R1}, \quad (1)$$

where the source's information symbol considers $s(t)$ with $\mathbb{E}\{|s(t)|^2\} = 1$ ($\mathbb{E}\{\cdot\}$ is expectation operation), P_s transmits the symbol to harvested energy and data processing at first hop; the channel estimation error is Δh that is random variables with $\sigma_{\Delta h}^2$.

The TSP protocol that is suggested in (Nasir et al. 2013), the energy harvesting at (R) can be computed as

$$E_h^{TSR} = d_1^{-m} \eta P_s \left(|h|^2 + \sigma_{\Delta h}^2 \right) \alpha T, \quad (2)$$

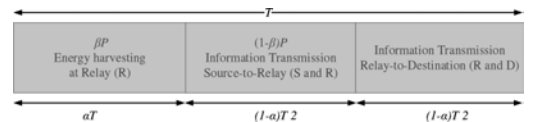


Figure 1. Energy harvesting protocol for relaying network.

where $0 < \eta < 1$ is the power conversion efficiency that rely on the harvested power circuitry and rectification process.

The PSR protocol that is suggested in (Nasir et al. 2013), the entire transmitted energy from source is divided into two components, βP used as harvested power function and $(1-\beta)P$ served for data processing.

Therefore, the energy harvesting is given as

$$E_h^{PSR} = d_1^{-m} \eta P_S \left(|h|^2 + \sigma_{\Delta h}^2 \right) \beta T. \quad (3)$$

Relying on the harvested power protocol, there is a tradeoff between the quantity of power transferred to (R) node and the standard of data transmission to (R)→(D) node.

In this study, we derive the overall harvested power protocol TPSR that is calculated by

$$E_h^{TPSR} = d_1^{-m} \eta P_S \left(|h|^2 + \sigma_{\Delta h}^2 \right) \alpha \beta T. \quad (4)$$

The average energy harvesting over Rayleigh fading channels that is based on (4), is computed as

$$E_{avg} = \mathbb{E}(E_h^{TPSR}) = \frac{1}{2} \frac{\eta P_S}{d_1^m} \alpha \beta (\Omega_h + \sigma_{\Delta h}^2). \quad (5)$$

Next, the transmitted energy from (R) node, P_R is calculated by

$$P_R = \frac{E_h^{TPSR}}{(1-\alpha)T/2} = \varphi P_S \left(|h|^2 + \sigma_{\Delta h}^2 \right), \quad (6)$$

where $\varphi = \frac{2\eta\alpha\beta}{d_1^m(1-\alpha)}$.

3 OUTAGE PROBABILITY AND THROUGHPUT ANALYSIS

Regarding AF relaying network, the signal first operations at (R) node and the received signal is amplified to send to (D) node.

The sampled signal at (R) in the first hop is given by

$$y_R(k) = \sqrt{d_1^{-m}(1-\beta)P_S} (h + \Delta h)s(k) + n^{(R)}, \quad (7)$$

where additive white Gaussian noises (AWGN) is $n^{(R)}$ ($\sigma_{n^{(R)}}^2$ is variance) that the average is 0. We have $x_R(t) = Gy_R(t)$ is the signal transmitted at (R), where the amplification factor is computed as

$$G^2 = \frac{d_1^m P_R}{(1-\beta)P_S(|h|^2 + \sigma_{\Delta h}^2) + d_1^m \sigma_{n^{(R)}}^2}. \quad (8)$$

The energy constraint factor from the energy of received signal can be obtained by (R) node. The sampled received at (D) is computed by

$$y_D(k) = \sqrt{d_2^{-m}} (g + \Delta g)x_R(k) + n^{(D)}, \quad (9)$$

where Δg is the channel estimation error with $\Delta g \sim CN(0, \sigma_{\Delta g}^2)$.

Replacing (7) into (9)

$$y_D = \sqrt{d_2^{-m}} G \underbrace{(Mhgs(k))}_{\text{Signal}} + \sqrt{d_2^{-m}} G \underbrace{\left(Mg\Delta hs(k) + Mh\Delta gs(k) + M\Delta h\Delta gs(k) \right)}_{\text{CSINoise}} + \sqrt{d_2^{-m}} G \underbrace{\left((g + \Delta g)n^{(R)} + n^{(D)} \right)}_{\text{AWGNNoise}}, \quad (10)$$

where $M = \sqrt{d_1^{-m}(1-\beta)P_S}$.

As a result, at destination, the received end-to-end SNR can be calculated as

$$SNR = \frac{|g|^2 |h|^2}{|g|^2 U_1 + |h|^2 U_2 + U_3}, \quad (11)$$

where $U_1 = \sigma_{\Delta h}^2 + \frac{d_1^m}{(1-\beta)P_S} \sigma_{n^{(R)}}^2$, $U_2 = \sigma_{\Delta g}^2$, $U_3 = \sigma_{\Delta g}^2 \sigma_{\Delta h}^2 + \frac{d_1^m}{(1-\beta)P_S} \sigma_{\Delta g}^2 \sigma_{n^{(R)}}^2 + \frac{(1-\alpha)d_1^m d_2^m \sigma_{n^{(D)}}^2}{2\eta\alpha\beta P_S}$.

3.1 Delay-limited transmission

In this case, it can be assessed, P_{out} , the outage possibility at a fixed transmission rate (R) (bits/sec/Hz), this paper can show the throughput where SNR_0 , the SNR_s threshold for intact obtained information at (D) is SNR_0 and $R = \log_2(1 + SNR_0)$.

The vital performance measure of transmission systems is $P_{out}(\cdot)$, which is outage probability and the predetermined threshold γ_{th} is larger than the received SNR. Therefore, the victorious transmission rate can compute by the outage probability. In the WPCN network, the outage probability $P_{out} = \Pr(SNR < \gamma_{th})$, can be given as

$$P_{out} = \Pr \left\{ \frac{|g|^2 |h|^2}{|g|^2 U_1 + |h|^2 U_2 + U_3} < \gamma_{th} \right\}, \quad (12)$$

where $\gamma_{th} = 2^R - 1$, and R (bits/sec/Hz) is the fixed source transmission rate. In the proposition 1, the analytical expression P_{out} can be extracted.

Proposition 1: The imperfect of CSI for the considered protocol at (D) node, the outage probability is computed by

$$P_{out} = 1 - e^{-\frac{U_1 \gamma_{ih}}{\Omega_g} - \frac{\gamma_{ih} U_2}{\Omega_h}} \Psi K_1(\Psi), \quad (13)$$

where $\Psi = \sqrt{4 \frac{\gamma_{ih}(U_3 + \gamma_{ih} U_1 U_2)}{\Omega_g \Omega_h}}$. The values of the exponential random variables $|h|^2$ and $|g|^2$ are Ω_h and Ω_g , respectively, and $K_1(\cdot)$ is the first order modified Bessel function of the second kind (Ryzhik & M 1980).

Proof:

The Cumulative Distribution Function (CDF) of $|g|^2$ which is the exponential random variable.

$$P_{out} \triangleq F(\gamma_{ih}) = \Pr \left\{ \frac{|g|^2 |h|^2}{|g|^2 U_1 + |h|^2 U_2 + U_3} < \gamma_{ih} \right\}. \quad (14)$$

The outage probability, P_{out} , is given by

$$P_{out} = 1 - \frac{1}{\Omega_h} e^{-\frac{\gamma_{ih} U_1}{\Omega_g}} \int_{\gamma_{ih} U_2}^{\infty} e^{-\left(\frac{y}{\Omega_h} + \frac{\gamma_{ih}(U_3 + \gamma_{ih} U_1 U_2)}{(y - \gamma_{ih} U_1) \Omega_g} \right)} dy. \quad (15)$$

Therefore, when SNR rises, the approximate outage is computed by

$$P_{out} = 1 - 2e^{-\frac{U_1 \gamma_{ih}}{\Omega_g} - \frac{\gamma_{ih} U_2}{\Omega_h}} \sqrt{\frac{\gamma_{ih}(U_3 + \gamma_{ih} U_1 U_2)}{\Omega_g \Omega_h}} \times K_1 \left(2 \sqrt{\frac{\gamma_{ih}(U_3 + \gamma_{ih} U_1 U_2)}{\Omega_g \Omega_h}} \right). \quad (16)$$

We use the formula obtains the last equality, $\int_0^{\infty} e^{-\frac{u}{4x} - vx} dx = \sqrt{\frac{u}{v}} K_1(\sqrt{uv})$, [(Ryzhik & M. 1980). 3.324.1]. This ends the proof for Proposition 1.

At (D) node, the throughput, τ is calculated by

$$\tau = R(1 - P_{out})(1 - \alpha)/2, \quad (17)$$

where the throughput in (17), relies on P_S , η , R , α , d_1 , and d_2 , $\sigma_{n|R}^2$, $\sigma_{n|D}^2$.

3.2 Delay-tolerant transmission

In this mode, the capacity of ergodic, C , complete the throughput at (D). It differs the delay-limited transmission mode, where (R) is the source transmits rate at fixed in order to meet a number of outage criteria, information can be sent at the rate which is lower than or equal to the capacity of the evaluated ergodic, C in this mode.

The ergodic C is computed by

$$C = E_{|h|^2, |g|^2} \left\{ \frac{1}{2} (\log_2(1 + \gamma)) \right\}, \quad (18)$$

where γ , relies on the random channel gains, h and g .

Proposition 2: The capacity of ergodic at (D) node is given by

$$C \approx \int_0^{\infty} \left(\frac{1}{\Omega_h \Omega_g} (\Omega_h U_1 + \Omega_g U_2) e^{-\frac{U_1 \gamma}{\Omega_g} - \frac{\gamma U_2}{\Omega_h}} \Psi K_1(\Psi) + \frac{2(2U_1 U_2 \gamma + U_3)}{\Omega_h} e^{-\frac{U_1 \gamma}{\Omega_g} - \frac{\gamma U_2}{\Omega_h}} K_0(\Psi) \right) \times (\log_2(1 + \gamma)) d\gamma \quad (19)$$

Proof:

To calculate the analytical expression for the capacity of ergodic, $f(\gamma)$ is the Probability Density Function (PDF) of SNR, which is evaluated first. The CDF can obtain the PDF, $F(\gamma)$ that is presented in Proposition 1.

Next, the ergodic capacity is given as below

$$C = \int_0^{\infty} f(\gamma) \log_2(1 + \gamma) d\gamma. \quad (20)$$

The PDF of SNR is given by

$$f(\gamma) = \frac{\partial F(\gamma)}{\partial \gamma} = \frac{\partial \left[1 - e^{-\frac{U_1 \gamma}{\Omega_g} - \frac{\gamma U_2}{\Omega_h}} \Psi K_1(\Psi) \right]}{\partial \gamma}. \quad (21)$$

For simplicity, $\Psi = \frac{4\gamma(U_3 + \gamma U_1 U_2)}{\Omega_g \Omega_h}$ are denoted in (21) can be rewritten by

$$f(\gamma) = \left[\frac{1}{\Omega_h \Omega_g} (\Omega_h U_1 + \Omega_g U_2) e^{-\frac{U_1 \gamma}{\Omega_g} - \frac{\gamma U_2}{\Omega_h}} \Psi K_1(\Psi) + \frac{2(2U_1 U_2 \gamma + U_3)}{\Omega_h} e^{-\frac{U_1 \gamma}{\Omega_g} - \frac{\gamma U_2}{\Omega_h}} K_0(\Psi) \right]. \quad (22)$$

Thus, the capacity of ergodic is rewritten by

$$C \approx \int_0^{\infty} \left(\frac{1}{\Omega_h \Omega_g} (\Omega_h U_1 + \Omega_g U_2) e^{-\frac{U_1 \gamma}{\Omega_g} - \frac{\gamma U_2}{\Omega_h}} \Psi K_1(\Psi) + \frac{2(2U_1 U_2 \gamma + U_3)}{\Omega_h} e^{-\frac{U_1 \gamma}{\Omega_g} - \frac{\gamma U_2}{\Omega_h}} K_0(\Psi) \right) \times (\log_2(1 + \gamma)) d\gamma \quad (23)$$

This ends the proof for Proposition 2.

The ergodic capacity C (bits/sec/Hz), equal to a fixed rate is being transmitted. The throughput at (D) node is given by

$$\tau = \frac{(1-\alpha)T/2}{T} C = (1-\alpha)C/2. \quad (24)$$

4 NUMERICAL RESULTS AND DISCUSSION

In this section, some illustrative samples depicts the provided transmission modes behavior of the outage probability and ergodic capacity. Particularly, the simulation for TPSR protocol is used in both delaytolerant and delay-limited transmission modes. Moreover, this paper can explore and justify the analytical throughput.

In the delay limited mode, we have the source transmission rate $R = 3$ (bits/sec/Hz), harvested energy efficiency $\eta = 1$, path loss exponent $m = 3$ and source transmission power, $P_s = 1$ (Joules/sec). The unit value normalized by the gap from the first hops to the second hop. In order to reduce complexity, similar noise variances at (R) and (D) node are denoted as $\sigma_{n[R]}^2 = \sigma_{n[D]}^2 = \sigma^2 = 0.01$. The value of the exponential random variables $|h|^2$ and $|g|^2$, are put to 1. The averaging expressions associated with end-to-end SNR and outage probability calculates by the experimental consequences and outage probability over 10^5 that is the random of Rayleigh fading channels g and h .

Fig. 2 depicts that the performance of system levels off in instance of ideal CSI. We can see that there are a downward trend in the throughput for imperfect CSI.

This tendency in Fig. 3 is upward trend when SNR rises, however, when SNR is low, the throughput is tiny.

The performance of throughput versus the harvested power time coefficients and energy splitting coefficients are investigated by Fig. 4 and Fig. 5. In this case, Monte-Carlo simulations for both transmission modes verifies and examines for the analytical results of throughput. Overall, the relaying network with ideal and imperfect CSI are enabled by the energy harvesting. It shows that the analytical and the simulation results match for both modes for all values of α and β . The figure shows that the throughput rises as α increases. Because when α is lower than the value of optimal α , less time for harvested energy. Hence, less power is harvested and due to higher outage probability, the lower level of throughput are at (D) node. On the other hand, when α is larger than optimal α , energy harvesting has more time and data

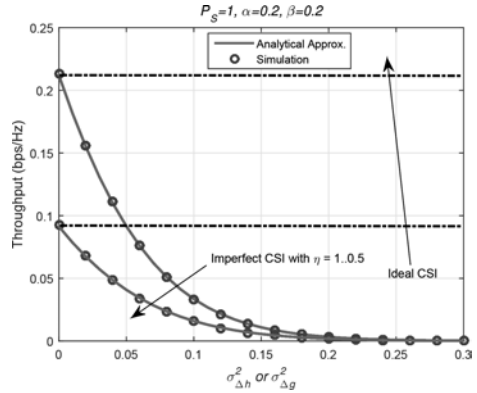


Figure 2. Comparison of throughput τ for imperfect CSI and Ideal CSI.

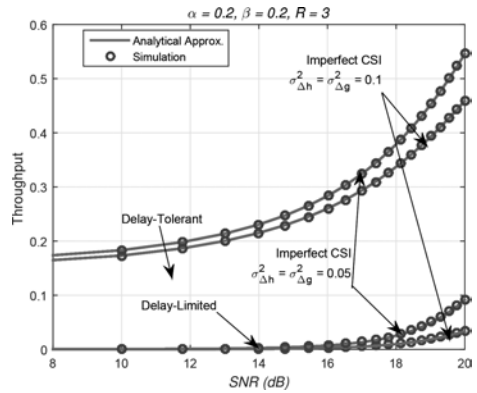


Figure 3. Throughput versus SNR for imperfect CSI $\sigma_{\Delta h}^2 = \sigma_{\Delta g}^2 = 0.1$ or $\sigma_{\Delta h}^2 = \sigma_{\Delta g}^2 = 0.05$

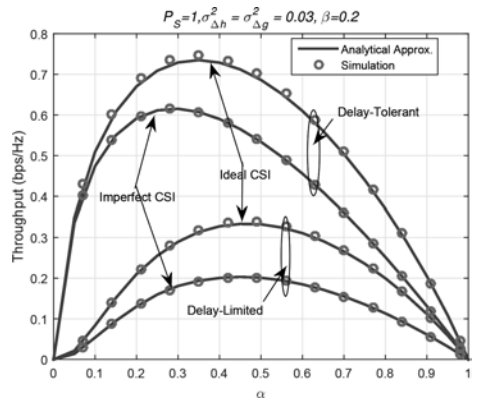


Figure 4. Throughput versus time switching factor.

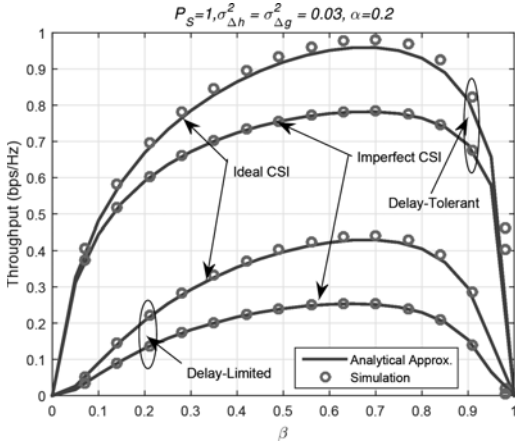


Figure 5. Throughput versus power splitting factor.

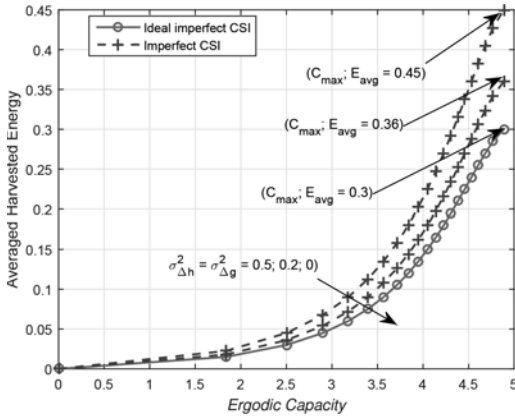


Figure 6. Averaged Harvested Energy for imperfect CSI with different $\sigma_{\Delta h}^2 = \sigma_{\Delta g}^2 = 0.5$ or $\Delta h = \Delta g = 0.2$ and Ideal CSI.

transmission has less time, so this results decline. We can see that both figure have similar tendency.

In Fig. 6, we illustrate the numerical results of energy harvesting tradeoff and ergodic capacity. We can be seen that the results of ergodic rises as the power harvesting declines. In addition, the ergodic capacity is less sensitive to channel estimation error than energy harvesting. Because of channel estimation error is linear with the level of power.

In the previous simulations, it is noted that this paper can calculate optimal energy factors and time in TPSR protocol when throughput is maximum (fixed $\alpha = 0.32$, $\beta = 0.58$, and $\alpha = 0.169$, $\beta = 0.69$). Evaluating the effects of noise on this results, we reveal in Fig. 7 the optimal throughput in two modes for various values of noise variance σ^2 at antenna. Another interesting point is that

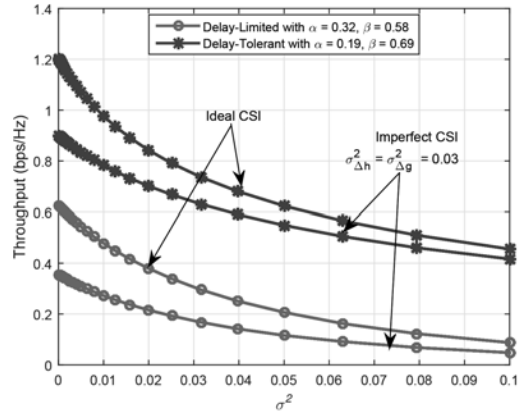


Figure 7. Throughput versus different of noise variance.

the performance of throughput gaps between two instances decline when rising noise variance.

5 CONCLUSION

In this study, the effect of imperfect CSI optimized and analyzed by the harvested power enabled AF relaying network. The optimal value can approach by investigating the duty of power-collecting receivers if the approximate energy splitting factors and time in harvested power are chosen. We show that, relying on the quantity of Channel State Information (CSI) and application scenario. The throughput result can modeled as a term of linear scale when noise variances, energy harvesting factors and SNR is changed. The simulated and analytical results depict that the power constraint relay with imperfect CSI is lower than ideal CSI, however, when the value of CSI error is small, the outage probability and throughput are remained.

REFERENCES

- Chalise, B.K., M. Wing-Kin, Y.D. Zhang, H.A. Suraweera, & M.G. Amin (2013). Optimum performance boundaries of ostbc based af-mimo relay system with energy harvesting receiver. in *Proc., IEEE Transactions on Signal Processing* 61(17), 4199–4213.
- De, S. & R. Singhal (2012). Toward uninterrupted operation of wireless sensor networks. *Computer* 45(9), 24–30.
- Grover, P. & A. Sahai (2010). Shannon meets tesla: Wireless information and power transfer. In *in Proc., IEEE International Symposium on Information Theory Proceedings (ISIT)*, pp. 2363–2367.
- Hyungsik, J. & Z. Rui (2014). Optimal resource allocation in full-duplex wireless-powered communication

- network. In *Proc. IEEE Transactions on Communications* 62(10), 3528–3540.
- Laneman, J.N., W.G.W. & D.N.C. Tse (2001). An efficient protocol for realizing cooperative diversity in wireless networks. *Proc. of IEEE Int. Symp. Inform. Theory (ISIT)*.
- Nasir, A.A., Z. Xiangyun, S. Durrani, & R.A. Kennedy (2013). Relaying protocols for wireless energy harvesting and information processing. in *Proc., IEEE Transactions on Wireless Communications* 12(7), 3622–3636.
- Ng, D.W.K., E.S. Lo, & R. Schober (2013). Wireless information and power transfer: Energy efficiency optimization in of dma systems. in *Proc., IEEE Transactions on Wireless Communications*. 12(12), 6352–6370.
- Rubio, J. & A. Pascual-Iserte. Simultaneous wireless information and power transfer in multiuser mimo systems. In *Global Communications Conference (GLOBECOM), 2013 IEEE*, pp. 2755–2760.
- Ryzhik, I.S.G. & I.M. (1980). Table of integrals, series, and products, 4th ed. academic press, inc.
- Tourki, K., Q.K.A. & M.S. Alouini (2013). Outage analysis for underlay cognitive networks using incremental regenerative relaying. *Vehicular Technology, IEEE Transactions*. 62, 721–734.
- Varshney, L.R. (2008). Transporting information and energy simultaneously. In *in Proc., IEEE International Symposium on Information Theory, ISIT*, pp. 1612–1616.
- Visser, H.J. & R.J.M. Vullers (2013). in proc., rf energy harvesting and transport for wireless sensor network applications: Principles and requirements. *Proceedings of the IEEE* 101(6), 1410–1423.
- Xiang, Z. & M. Tao (2012). Robust beamforming for wireless information and power transmission. *Signal Processing Letters, IEEE*. 1(4), 372375.
- Zhiguo, D., S.M. Perlaza, I. Esnaola, & H.V. Poor (2014). Power allocation strategies in energy harvesting wireless cooperative networks. *IEEE Transactions on Wireless Communications* 13(2), 846–860.
- Zhou, X., R. Zhang, & C.K. Ho (2013). Wireless information and power transfer: Architecture design and rate-energy tradeoff. *IEEE Transactions on Communications* 61(11), 4754–4767.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Content based video retrieval—towards full multimedia join

R.A. Abinader

Antonine University, Beirut, Lebanon

P.G. Gedeon

Notre Dame University, Louaize, Lebanon

ABSTRACT: Content Based Image Retrieval is the appliance of various pixel-level computation techniques, which assists the connection between image querying and human perception. Low level features are extracted by CBIR systems like color, texture shape and position. Next to audio features, a gap is formed between these content based low-level ones and the high-level semantic concepts. Multimedia Join comes to build the connecting bridge over these breaches, one more improvement on the human-machine interface. Video is an important brick in Multimedia Join since it's a complex type of data. Images on the other hand are the building blocks of videos. In this paper we introduce, develop and test a framework of video join, by applying content based image retrieval methods. A multimedia join model was conceived and partially implemented on the video level. A new approach is proposed and implemented in order to optimize the joining time. The testing phase major lines consisted on video-to-video association and on scene-to-scene association.

Keywords: Multimedia Join, Image, Video, Content Based Retrieval, Query by Example, MPEG

1 INTRODUCTION

1.1 “A picture is worth ten thousand words” *Confucius*

Most video search engines rely on keywords. Depending on how powerful the retrieval engine is, the search can go from the simplest task (like searching video titles), to more complex tasks (like searching the names of actors, producers etc. that relates to a certain video). This textual type of queries is essential; however, it is not sufficient. Numerous are the videos that are not accompanied by textual meta-data, many are even lacking any relevant information (like web-based video outlets where anyone can post personal videos such as YouTube). And when it is not the case, the videos' descriptions are written only in one language, hence the reference to the semantic gap¹.

On another hand, a report recently released by Adobe Corporation (2014)² suggests that Online media and TV streaming rises 388 percent yearly and that “more consumers than ever are turning to web-based media outlets in order to watch television, rather than relying upon traditional cable providers”. In the midst of these facts comes the questioning about online video usage, five years from now.

Therefore, the need for systems capable of handling video queries by example is increasing with each video published on the internet.

Consequently, the “video query by example” imposes itself in many application scenarios. For example, in education this concept allows to realize and produce computer-based training courses; It can also reference books like encyclopedias and almanacs. As a result of multimedia join, a user can go through a series of presentations, text about a particular topic, and associated illustrations in various information formats.

Multimedia join consists of making possible, the passage between one type of media to another. For example, making a textual query on the core of DVDs, or making a video clip query on several audio recordings, etc.

Since video is a wide-ranging media type, because it contains images, sounds and texts, it becomes evident to start working on video joins. This subject will be covered in this paper by studying various video comparison and retrieval techniques in the below paragraphs 1.2, 1.3 and 1.4.

1.2 Property based retrieval

Property Based Retrieval consists of extracting the video's properties, during the insertion in the data-

base, and then storing them as metadata correlated to the video sequence. These properties include: format, width, height, video duration.

Even though this is a simple approach, it is rarely offered to web consumers. One of its usages, is the mime type property, which depicts the video format such as “.avi” or “.wmv” etc. A concrete example of this procedure is found on the Google search engine, on which, the consumer can define the file type of the video he is looking for, i.e. by adding to his query “filetype:wmv”.

1.3 Mental image search

Boujemaa & Fauqueur (2006)³ in their work on Mental Image Search, state that this technique is based on a procedure where the color regions of an image are detected and retrieved. Every region is cropped; then, the images and their regions are stored in a database. After that, a thesaurus linking all these regions is constructed.

During retrieval, the user provides an assembled image that contains only color regions proposed by the system itself.

Though this method is only employed to search for images, its application to video retrieval was not studied.

1.4 Spatio-temporal search

Mehmet & Özgür & Ugur (2003)⁴ state that spatio-temporal search is composed of two separate processes: One that manages spatial data and another one that manages temporal data.

Though this approach provides full support for object trajectories it neglects simpler video modifications. For example, the search for a sequence that contains a certain surface texture like water or sand will not be a success because of the fact that texture variations are dropped.

2 THE MULTIMEDIA MULTI-LAYERED MODEL, 3ML

2.1 Introduction

A model—the 3ML, or the Multimedia Multi-Layered Model—is proposed in this paper as a framework for multimedia management and integration.

The model is capable of managing any multimedia category defined as being: image, audio, video, document, XML files, etc.

Video sequences present all these categories. A DVD for example proposes sound, images (since the visual data is made of series of images) and text data (subtitles, meta-data etc.). The proposed

3ML framework integrates the multimedia join principles in order to preserve the unity of these types all over the layers.

Multimedia join is as such defined by being the mechanism of linking several types of media, under the purpose of making it a gateway between one form and another; thus taking the automated information’s definition to the next level, where this definition will not be bounded by text only nor by video only nor by sound only. These media entities, and by having tight-coupled connections will integrate into one but multi-sided definition.

2.2 Implementation

The model represents three major layers: image layer, sound layer and text layer. Each one of these layers will make possible the connection between data of the same type. Then, logic connections relate each layer with the other. On this version of this model the logic relations are: “or”, “and” and “distinct”.

2.2.1 Image layer

On this layer, the bond between graphical types of data occurs. So, images are taken from video types, represented by “Video.Image”, and still images are considered as one entity. Since audio data does not represent any graphical information, the audio side, in the “Image” plan of the image below, is blocked.

2.2.2 Sound layer

On this layer, the bond between sound types of data exists. So, sounds (music, speech etc.) are taken from video types, represented by “Video.

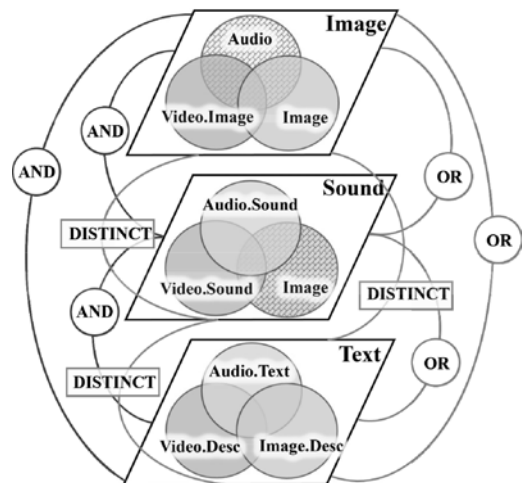


Figure 1. The multimedia multi-layered model, 3ML.

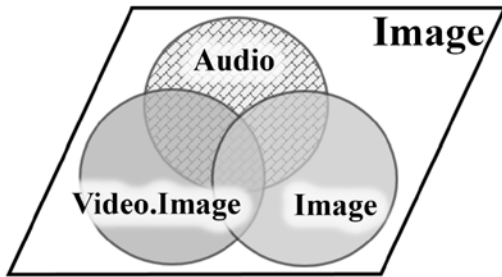


Figure 2. The image plan in the multi-layered model, 3ML.

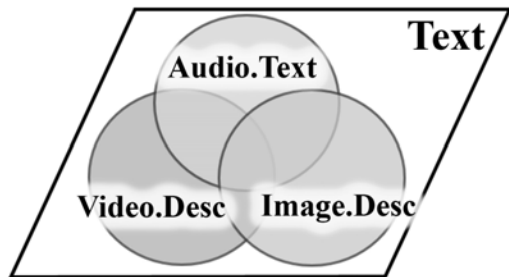


Figure 4. The text plan in the multimedia multi-layered model, 3ML.

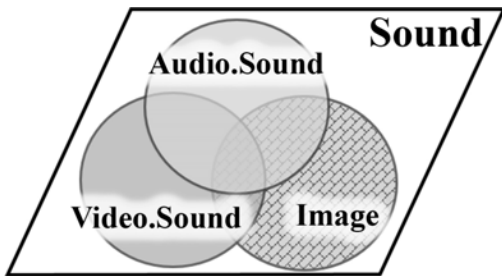


Figure 3. The sound plan in the multimedia multi-layered model, 3ML.

Sound”, and audio sequences are considered as an entity. Since still image data does not represent any audio information, the image side in the “Sound” plan of the image below is blocked.

2.2.3 Text layer

On this layer, the bond between text types of data exists. This data could be plain text description or even structured like XML etc. So, texts (subtitles, metadata, descriptions, speech that is converted to text in videos that do not have subtitles, or texts parsed from any external audio file, etc.) are taken from video types (in “Video.Text”), from Audio (in “Audio.Text”) and from Image’s descriptions (“Image.Desc”).

2.3 Application scenario

Many application scenarios come in handy. One particular is when one would want to watch all the films where [the aria “Una Furtiva Lacrima” is sung and the famous opera singer Luciano Pavarotti acts] or [where Plácido Domingo appears]. This search becomes very tricky when the search performer does not know the air’s name and only has a German sample of it; moreover, he only has a photo of Domingo.

The solution to this scenario, and according to the 3ML, comes with the following: “Una Furtiva Lacrima” is treated in the second layer (rhythm, harmony, pitch, etc.), the videos returned are filtered with the results of the third layer where “Luciano Pavarotti” is queried (via textual search in the metadata); the resulting films will be formulated this way:

[FILMS(Audio layer 2 AND Text layer 3) + FILMS(Image layer 1)]

Nevertheless, hiding this advanced technology behind a user-friendly interface remains another challenge.

In conclusion, image is the common good example to all videos; whether the videos’ spoken language is different, or whether videos do not have any description attached to them as it is the case with the majority of personal videos on the internet.

The picture resides the silent universal type to all videos. From this point we find it of a major importance to work on the first layer of the 3ML, by applying CBIR techniques on video joins.

3 CONTENT-BASED IMAGE RETRIEVAL, CBIR

3.1 Introduction

Content-Based Image Retrieval, CBIR is the process of retrieving one or many images from a repository on the basis of very specific features that are extracted from the images themselves⁵.

A CBIR based engine is a powerful image-based search engine, that considers these human perceptual features: color, shape, texture and location. CBIR returns images similar to a query image, providing more realistic approaches closer to the human way of seeing.

Edward & Li (2001)⁶ define the color parameter in content based image retrieval as computably defined by the three variables: red, green and blue.



Figure 5. Two identically shaped houses.

Since it is possible to obtain more than 32 bits of color by using all the combinations; scientists have divided the colors in two systems JND and JNS (respectively Just Noticeable Difference and Just Not the Same). By this procedure the whole color spectrum is bounded by eleven colors only.

The downwards example is presented to show practically how CBIR compares two images:

The comparison technique depicts a function that has at least three parameters:

- One: an image signature generated by processing the first image “SigHouse1”
- Two: an image signature generated by processing the second image “SigHouse2”
- Three: a number from 0 to 1 concerning each of the following: color, shape, texture and location

The function returns a percentage of similarity between the two houses. For example:

Function Compare (Sig_House1, Sig_House2, Color = 0.3, Shape = 0.7) = 65%.

Though the technique above is integrated in many Database Management Systems, it remains limited to managing still images and does not have any extension to work on video types.

3.2 Matching—Frame analyzer module

On another level, and according to the Motion Picture Experts Group⁷, MPEG standards, four types of frames exist in MPEG coded videos: I-Frames that are quality pictures, P-Frames, B-Frames, D-Frames.

Since I-Frames offer more significant and accurate data, a module capable of extracting these frames was required to be implemented.

The integration of the frame analyzer with a CBIR system is presented in the next part.

The proceedings are explained in the following example:

The top filmstrip, in the image above, represents the images of the video sample query after frame analysis has been applied.

The lower filmstrip represents the images of a certain video in the database. The objective is to compare these two videos.

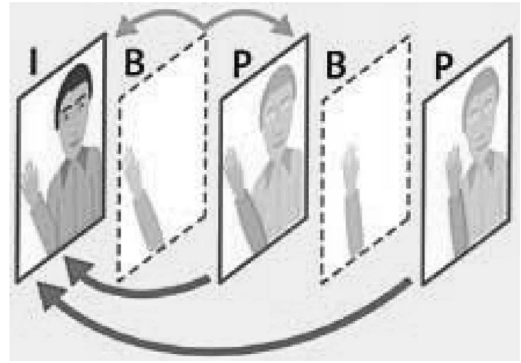


Figure 6. The I, P and B frames.



Figure 7. An example of the frame analyzer.

Another logical representation of the filmstrips is showed in the vectors below. For each image one image signature object is generated by the CBIR system. The below contains F plus a number, and Sig plus a number.

We mean by this representation that for each I-Frame image retrieved:

- F is the index of the image from each time interval
- Sig is the image signature related to each image after being processed by the CBIR engine.

3.3 Retrieval method 1—Active Signature (AS)

When retrieval is started, this procedure will work like the following:

- We compare the first signature of the first image (retrieved from the I-Frame of the example video) with the signature of the first image (of any video in the database, and which is not the example video itself).
- Once the first signature is compared with all other signatures, the maximum similarity percentage found is stored in a vector. If a positive relaxation number is defined, the procedure stores all the values, that occurred in the following scene: [Max(Similarity); Max(Similarity) – Relaxation]
- Then moves on to the next signature (Sig 2, etc.)

F1	F2	F3	...	Fn
Sig1	Sig2	Sig3	...	Sign

F'1	F'2	F'3	F'4	F'5	...	F'm
Sig'1	Sig'2	Sig'3	Sig'4	Sig'5	...	Sig'm

Figure 8. The frames being transformed to signatures.

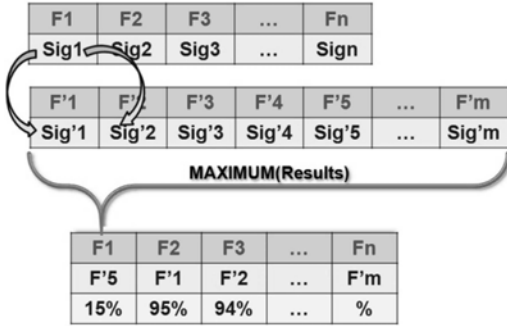


Figure 9. The signature being transformed to percentages.

3.4 Retrieval method 2—Passive Signature (PS)

3.4.1 Introduction

The idea is initiated after the recognition of the following facts:

- A video resembles the best to itself.
- Like the spatial approach, projections on the (X,Y) axes were made to simplify calculus. In this type of study, the major constraint is the retrieval time since, to compare only two videos, with ten I-Frames each; the system has to make one thousand signature comparison.
- Working in an axis system composed of sub-objects like Color, Texture, Shape and Location (which are subs of the generic Image-Signature object) is slower than working in a numeral space. In other terms, comparing two Image-Signature objects takes more time than comparing two numbers like 78 and 90.

In the following image we represent both the signature space and the numerical space, and how each axis in the image signature is correlating with the numerical one.

3.4.2 Passive signature procedure

Since the retrieval time during the search process is more important than the insertion time of videos, a second method is proposed and named the



Figure 10. Both spaces, the signature space and the numerical one.

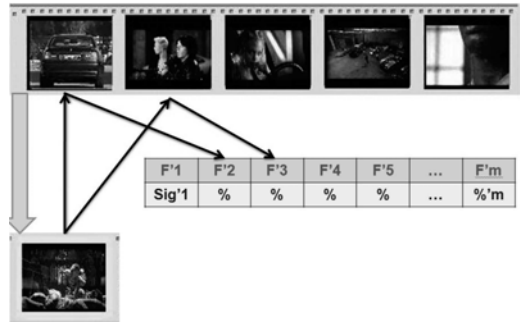


Figure 11. The passive signature procedure.

“Passive Signature” procedure. It is named as such because the signature is used during the insertion time then neglected during the retrieval time. This method follows these steps:

- We define one image for each video to be what is called “main image”.
- The “main image” will be affected with a “salient” signature, and all other images with the “minor” one.
- Then, we compare each minor image signature, with the salient image one.

All the above happens during insertion time and a vector is created to shadow the comparison results.

Then, and when retrieval is triggered:

- A stored procedure will compare each signature of the processed video with the Main Signature of the example video only.
- Then, this procedure will store the percentage of this comparison in temporary columns for each time frame.
- Afterwards, this procedure will work like the Active Signature but, instead of comparing sub-objects, it will compare real numbers; thus, dramatically decreasing the retrieval time.

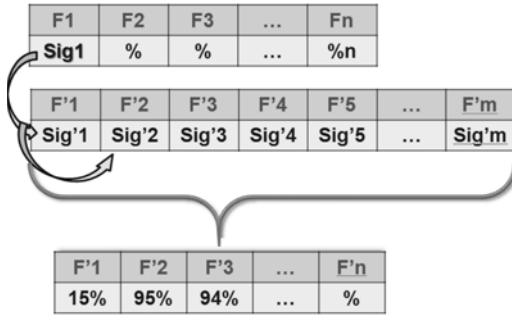


Figure 12. The pre-processing during retrieval.

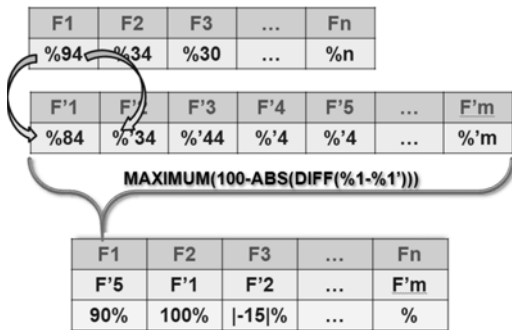


Figure 13. The processing during retrieval.

3.5 Search types

Two result interpretations of the search methods were conceived: the search by Video and the search by Scene.

The join directions thought of are: One Left to Left, One Left to Right, One Right to Right, One Right to Left, Many Left to Left, Many Left to Right, Many Right to Right and Many Right to Left.

3.5.1 Join by video

This retrieval will propose video results similar to one or more query videos entered. Moreover, whole video databases can be joined together.

For the join process to succeed, the user has to define the color, shape, texture, location and relaxation parameters for a query of type Active Signature; Or, only the relaxation parameter for the Passive Signature query. Color, texture, shape and location intervals are [0; 1]; and the relaxation parameter is a percentage.

We anticipated four ways to interpret the join results:

- Maximum: shows the best videos that match.
- Average: shows the videos relatively similar.



Figure 14. The prototype main window.



Figure 15. The retrieved results by video search.

- Minimum: shows the videos that were found dissimilar.
- Relaxation: shows the best videos that match in function of the number of occurrences.

3.5.2 Join by scene

This retrieval will propose scene by scene results. As for the Join by Video, the user has to define the color, shape, texture, location and relaxation parameters for a query of type Active-Signature; Or only the relaxation parameter for the Passive Signature query.

For each registered scene, of the query video (displayed in the list of the left), corresponds the scenes (displayed in the list of the right). That gives a more detailed aspect than the by Video retrieval.

Each item of the list on the right contains: the corresponding time interval (in seconds), the percentage of similarity and the corresponding video index. Note, that the name of the video in question is displayed above the list, in this case Simpsons_MV.avi.

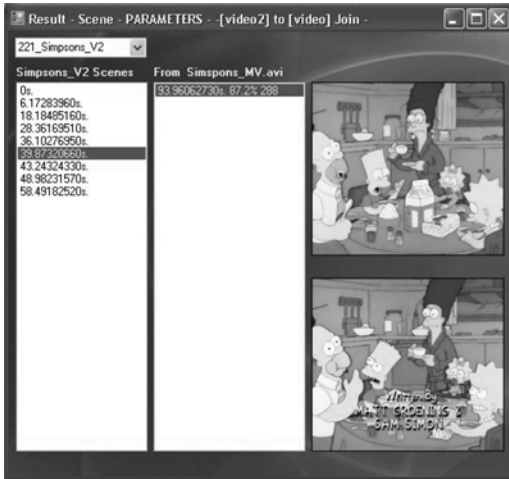


Figure 16. The join by scene window.

When a certain time frame is clicked, the corresponding scene is previewed in the display space of the right. The upper picture corresponds to the list of the right.

When the relaxation parameter is positive, several records are more likely to be displayed.

The screen where the search happens is shown in the image below.

3.6 Tools

3.6.1 Oracle 10g R2 database and interMedia component

Oracle interMedia enables Oracle 10 g to manage image and other multimedia in an integrated style with other types of information that are necessary to any project's evolution: audio, video, text, etc. Additionally, storage costs are reduced since managing image data that is stored with other traditional types of data is easier. Consequently, development ends up by being more intuitive and spontaneous.

Oracle Corporation's benchmark⁸ proves that storing images, in Oracle interMedia datatypes, in the Oracle Database 10 g simplifies application architecture and has no measurable impact on the storage and retrieval performance of the applications.

3.6.2 Microsoft .Net—Visual C#

Microsoft .Net Visual C# Language⁹ is a very interoperable language since it includes native support for the COM and windows based applications allowing restricted use of native pointers.

C# allows the users to use pointers as unsafe code¹⁰ blocks to manipulate unmanaged code guarantying more modules implementation as well as high performance when this one is needed.

Since we are manipulating large amount of data, a video can reach more than 30GB, C# was selected because of its support to native pointers in the .Net framework.

3.6.3 Data model

The database schema conceived is the following:

- Two tables video: for storing videos and later to retrieve them.
- Two tables image: for storing the extracted images and their time of occurrence inside the video.
- One table comparison: a temporary table where the retrieval results are stored. This table is directly connected to the result's interpreter of the main application.

3.7 Prototype

3.7.1 Retrieval performance

For the retrieval process to be the fastest possible, and more important to make possible future evolutions of the prototype (like cloud computing, mobile application integration, etc.), the whole execution is held by stored procedures on the DBMS level; only indexes are passed to the front-end application for interpretation.

The tests were done on a computer system having the following configuration:

- The DBMS was mounted on the same computer as the application.
- The total images of the DB: 600
- The total videos: 13
- The system's processor: Intel Pentium M 1.7 GHz
- The number of Windows XP processes during the tests: 70

The number of images is relatively large to the number of the video sequences. This factor was made on purpose in order to create an overload during the tests.

The purpose of these tests is to prove the processing time difference between the Active Signature method and the Passive Signature one. The change of the parameters (color, shape, texture and location) does not have much effect on the processing or benchmarking time, so the testing was executed on the color parameter where the color value was set to one.

In the below paragraphs 3.7.2 and 3.7.3 we show the test results in a Join by Video search and in a Join by Scene search where:

- The head of each column represents the number of images of the query video.
- The second row represents the processing time (in seconds) for the Active Signature method.
- The third row represents the processing time (in seconds) for the proposed color procedure in the Passive Signature.

3.7.2 Join by video

Registered scenes	9	18	45	170
Active Sig. (s)	11.6	19.5	35.9	110
Passive Sig. (s)	2.5	3.6	8.7	36

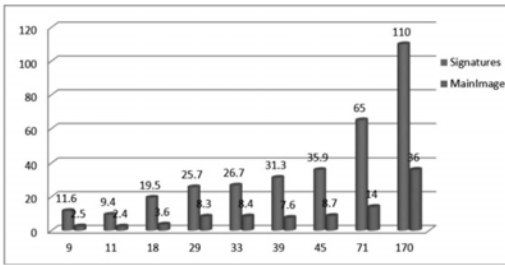


Figure 17. The join by video via the active signature method versus the passive signature method.

As the result above shows, when the number of images reaches 170, the processing time hits 110 seconds for the active signature method versus 36 seconds for the passive signature one.

3.7.3 Join by scene

Registered scenes	9	18	45	170
Active Sig. (s)	9.3	15.8	27.2	75
Passive Sig. (s)	1.5	1.3	2.1	5.5

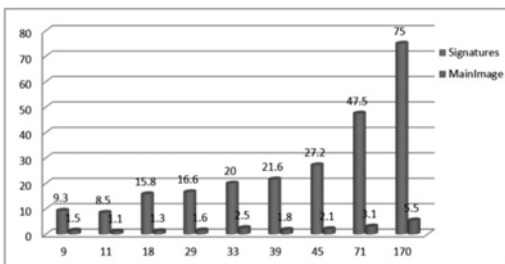


Figure 18. The join by scene via the active signature method versus the passive signature method.

As the result above shows, when the number of images reaches 170, the processing time hits 75 seconds for the active signature method versus 5.5 seconds for the passive signature one.

4 DISCUSSION

The results presented in this paper stand very promising when it comes to video retrieval using content based image retrieval techniques; whether in accuracy through the Active Signature method, or in speed through the Passive Signature method.

Moreover, the prototype was conceived and developed using sharp software engineering practices; this led to extending the human perceptual features (color, texture, shape, location) by enriching them with more attributes such as a “relaxation” factor, logical operators such as “and” “or” and “distinct”, a scene and video matching option, and others.

Comparing to other techniques such as the mental image search², the proposed method’s scope is wider since it spreads to cover video retrievals and is not limited to image ones.

On the other hand, the spatio-temporal search³ lacks behind the proposed approach since it neglects simple video modifications such as texture.

However, the property based retrieval technique could be easily integrated in the 3ML, the proposed model, which considers multimedia metadata as being an additional filter attribute.

In all cases, other researcher’s techniques can be hypothetically integrated in the architecture of the prototype which was designed to be modular based. This is made possible by assigning a retrieval weight for each technique and then, aggregating subsequent results into a complete one.

During all the phases of this work the idea of integrating different systems was more present than replacing other techniques.

At the end, the 3ML’s full potential is unlocked when video retrieval is combined with text retrieval, and the latter one combined with audio retrieval. Then, all the logical operators figuring in the model such as “and”, “or” and “distinct” come in hand to deliver one comprehensive system capable of handling intuitive, fast and accurate user queries on complete unknown multimedia objects.

5 CONCLUSION AND FUTURE WORK

We have presented a study about Content Based Video Retrieval, CBIR. The system implemented was based on a proposed model, the 3ML, which aims to stratify each type of media in its appropriate layer; thus making possible multimedia joins.

By this, the problem was transformed to be a problem of joining video data types.

Then we proceeded by implementing CBIR mechanisms in a prototype, to deal with the image layer of the 3ML.

Two retrieval methods were proposed: by-video and by-scene. Then, each of these methods was developed using two techniques: Active and Passive Signature.

The difference in the retrieval time was very noticeable between the two approaches. The time of the Passive Signature was only 7% of the Active Signature one, in cases where the video query enclosed 170 scenes.

Moreover, it was noticed that the processing time is bounded to:

- The video format (avi, mpeg, etc.) during the scene analysis processing. Though the prototype accepts several video formats, it is essential that the data be encoded as implied by the MPEG-1 or MPEG-2 standards.
- The image size, resolution, during the insert and retrieval processing.

Though the correctness of the Passive Signature method was noticed to be lower in certain cases than the Active Signature one, it made us think of several future works.

Interesting possibilities, yet unexplored, are to investigate the Passive Signature method boundaries in other challenging scenarios.

The first possibility is trying to set one main image for more than one video, so, when new videos are inserted in the system, they are directly compared to this one main image; thus decreasing spectacularly the retrieval time.

The second possibility we could think of, is that in order to achieve a very small factor Processing

Time over Correctness, we should explore a hybrid solution Active/Passive solution. For example, for each video, divide the scenes to two or many groups: one to be compared with the Active Signature method, and one group to be compared with the Passive method one.

REFERENCES

- [1] Harea, J. & Lewisa, P. & Enserb, P. & Sandomb, C. 2006. Mind the Gap: Another look at the problem of the semantic gap in image retrieval.
- [2] Adobe Corporation, *Video Benchmark Report 2014*, <http://www.zdnet.com/article/online-media-tv-streaming-rises-388-percent-yearly-report/>
- [3] Boujema, N. & Fauqueur, J. 2006. Mental image search by boolean composition of region categories.
- [4] Mehmet, E. D., Özgür U., Ugur G. 2003. Rule-Based Spatio-Temporal Query Processing for Video Databases.
- [5] Gudivada, V. & Raghavan, V. 1995. Content Based Image Retrieval Systems.
- [6] Edward Y. & Li C. & Li B. 2001. Toward Perception-Based Image.
- [7] Moving Picture Experts Group MPEG, <http://www.mpeg.org/>
- [8] Oracle Corporation, Using oracle Intermedia in Retail Banking Payment Systems, <http://www.oracle.com/technetwork/testcontent/imedia-in-bank-pay-sys-1-133627.pdf>.
- [9] Microsoft Corporation 2003. Visual C# Language, [https://msdn.microsoft.com/en-us/library/aa287558\(v=vs.71\).aspx](https://msdn.microsoft.com/en-us/library/aa287558(v=vs.71).aspx).
- [10] Microsoft Developer Network 2015. Unsafe Code and Pointers, <https://msdn.microsoft.com/en-us/library/t2yzs44b.aspx>.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Modeling the seller's strategy when sales depend on buzz

Olivier Lefebvre

Olivier Lefebvre Consultant, Paris, France

ABSTRACT: The goal of this paper is to compare the efficiency of two selling methods, “perfect information” and buzz. We model the buzz supposing that the consumers are “lazy”. That is to say, they pay attention to the buzz a single time, and are informed about one product, when two products are sold. Then we use game theory to compare the outcomes when two selling methods are chosen, “perfect information” (advertising, marketing) or buzz. Standard goods are concerned. We demonstrate that there is a single Nash equilibrium when two competitors sell goods using buzz. As a single Nash equilibrium exists also when “perfect information” is chosen (Bertrand competition) the comparison is possible. It appears that the advantages of buzz are dubious (as a selling method). The reason is simple: there are customers lacking.

This model also allows describing a particular kind of competition. It is when two competitors sell experience—based goods thanks to buzz. Videogames are an example. We describe the case of the first version of the game delivered for free.

Keywords: economic competition, game theory, consumer's behavior, videogame industry

1 INTRODUCTION

We start from an anecdote told by Raymonde Moulin in her book “The market of painting in France”. In Paris between the two World Wars there were two famous art dealers, Kahnweiler and Wildenstein (the both sold the paintings of young Picasso). Their strategies were very different:

Kahnweiler bought paintings from young talented painters at a low price, then sold them quickly to trigger buzz and make the painter renowned. Then the price of the paintings of this painter increased, and again he bought and sold his paintings...

Wildenstein also bought paintings from young painters who were not yet renowned, then stored them, selling them perhaps after twenty years when the prices have very much increased. His motto was: “boldness at the time of purchase, patience at the time of sale” [4].

This anecdote shows that a strategy relying on buzz can succeed.

Therefore to compare the efficiency of the two selling methods, “perfect information” and buzz, is interesting.

But before presenting the framework of the paper we have to define what “perfect information” and buzz are. Also we have to give details on the kinds of goods which are concerned.

1.1 Defining “perfect information” and buzz

“Perfect information” means advertising campaigns which allow the consumers knowing the existence of

the products (services), their main characteristics and their utility. When each consumer knows the utility of each product sold, he (she) buys the product providing the higher net utility (the difference between the utility and the price) or he (she) does not buy any product if all the net utilities are negative.

The buzz is merely the word to mouth. But it depends on the mindset of the consumers: they can be curious, have a mindset focused on research, or be “lazy”. Suppose there are two products in competition, A and B. If the consumers are curious, when the buzz about A (B) reaches one of them he will seek information on B (A). He will be informed about the two products. If the consumers are lazy, when one of them is reached by the buzz about A (B) he pays attention to the product A (B) only. Perhaps he buys A (B) or not. If he does not buy A (B) he is no more interested in the buzz about these products. This behavior is not in accordance with his interest. It is the consequence of his laziness. In the words of Kahneman this consumer uses his System 1 (intuition, problems easy to solve) and not his System 2 (logical reasoning). In his book “Thinking fast and slow” Kahneman studies laziness. We quote him: “Laziness is deeply embedded in our nature”. He speaks also of “cognitive ease”. The behavior of the lazy consumer who is disappointed (he has been reached by the buzz concerning a product and did not buy it) and is no more interested by any product of this kind, is explained by the “anchoring effect”. An “anchor” is a first judgment which influences definitively the ideas on some topic [1].

Clearly, the comparison which makes sense is between perfect information and buzz, the

consumers being lazy (if they are curious the outcome is the same).

1.2 What kind of good is concerned?

There are three kinds of goods:

Trust based goods require this condition: the honesty of the seller is well known. Examples are a solicitor, an arbitrator or a broker. The only strategy is to become renowned thanks to buzz.

Experience based goods require to have been tested by the consumer, before he can know the utility. Examples are some medical treatments and videogames. The only strategy possible for sellers is buzz.

Standard goods allow consumers knowing easily their utility as soon some information is available. There are many examples (cars, tours, houses etc.). Only these goods allow a comparison between the two selling methods, perfect information and buzz.

Now we can describe the framework of this paper. We shall consider four cases:

Case 1. We resume the story of Kahnweiler and Wildenstein. It is to show thanks to a simple imagined example that the strategy which banks on buzz (that of Kahnweiler) can allow winning more money than the one neglecting buzz (that of Wildenstein).

Case 2. Here we consider a single seller using buzz. It allows modeling the buzz in a simple way. Also, we state that the seller will decrease his price (compared to the case of perfect information) to win time. It is to accelerate the buzz. The consumers are winning.

Case 3. Here two sellers in competition use buzz, but the consumers are supposed curious. The outcome is the same than when perfect information is chosen. We shall use game theory and justify this choice. The case 3 allows us presenting what is called Bertrand competition in game theory.

Case 4. Two competitors sell their products using buzz, the consumers being supposed to be lazy. To model the buzz leads to an equation of the Volterra Lotka kind. We demonstrate that one Nash equilibrium, at least, exists.¹

The formulas giving the two profits are:

$$P_1 = N (p_1 - c_1) D_1 (p_1, 1) x_0 I^{D_1(p_1, 1)}$$

$$P_2 = N (p_2 - c_2) D_2 (1, p_2) y_0 I^{D_2(1, p_2)},$$

where the p_i and the c_i are the prices and the costs, N is the total number of consumers, Nx_0 and Ny_0

¹It is in pure strategies (which are not probabilistic). Otherwise there is always one Nash equilibrium, at least, if mixed strategies are possible (it is the Kakutani theorem).

are the numbers of informed consumers at the start ("sowing") and D_1 and D_2 are the demands, functions of p_1 and p_2 (considered as probabilities). The quantity I is given by:

$$x_0 I^{D_1(p_1, 1)} + y_0 I^{D_2(1, p_2)} = 1.$$

In the general case the calculations are too complex to state that this equilibrium is the single one. But in the symmetrical case (that we shall define) it is simpler: one can demonstrate that there is a single Nash equilibrium. This allows the comparison between the two Nash equilibriums, when perfect information is used and when buzz is used. It appears that the advantages of buzz are uncertain. It is not that the prices will increase very much. But one of the two parties (the sellers and the consumers) will lose. If the prices increase, the consequence is that the consumers' surplus decreases. If the prices decrease, the consequence is that the profits of the sellers decrease. Possibly the two parties lose, but one at least loses.

Our conclusion is that we shall live during a long time in a world where there are experts in advertising and marketing. When standard goods are concerned, either the sellers either the consumers benefit from "perfect information" (compared to buzz). Of course, the definition of buzz is somewhat arbitrary (the "laziness" of the consumers). But it gives an idea of the consequences of laziness of the consumers when buzz is used. Advertising is intrusive but it obliges the consumers to pay attention to the characteristics of the products sold. If one fears too much power for the experts in advertising, the solution is consumerism: associations give advices to consumers on the quality of the products, their characteristics, their prices.

2 CASE 1: THE STORY OF KAHNWEILER AND WILDENSTEIN

We imagine two phases: during the phase 1 the price of the paintings increases from 100 to 150 and during the phase 2 it increases from 150 to 200. Kahnweiler buys more paintings because he makes the prices increase and the painter works

Table 1. The strategies of Kahnweiler and Wildenstein.

	Phase 1	Phase 2
Prices	100 → 150	150 → 200
Kahnweiler	Buys 100 paintings at the price 100 and sells them at the price 150	Buys 120 paintings at the price 150 and sells them at the price 200
Wildenstein	Buys 100 paintings at the price 100 and sells them at the price 200	

more. Of course he sells the paintings to trigger buzz and make the prices increase. In the example Kahnweiler wins more money (11000) than Wildenstein (10000). The two strategies are shown in the Table 1.

3 CASE 2: A SINGLE FIRM SELLS ITS PRODUCT USING BUZZ

A firm sells its product at a price p and has a cost c . Its profit is $P(p) = (p - c) D(p)$ where $D(p)$ is the function of demand, decreasing and concave. At the instant t there are $N \times (t)$ consumers reached by the buzz (N is the total number of consumers). The number of consumers having bought the product is $x(t) D(p)$. They are “active agents” of the diffusion of the buzz. Therefore:

$$N dx = k_0 \times (t) D(p) N [1 - x] dt.$$

The number of consumers newly informed about the product (during the time dt) is some proportion of the encounters between those informed and having bought the product and those not informed. These encounters are the efficient encounters. Possibly the constant k_0 encapsulates the use of social networks and Internet (there are more efficient encounters). Finally:

$$x = 1/1 + Ke^{-kt}, \text{ with } K = 1 - x_0/x_0 \text{ and } k = k_0 D(p). \quad (1)$$

At the start there is a small number $N x_0$ of informed consumers (it is a kind of “sowing”). When $t \rightarrow \infty, x \rightarrow 1$. After time enough the value of the profit is $P(p)$. The seller has chosen $p_m = \text{Max}_p P(p)$.

There is no difference between the outcomes of perfect information and buzz.

But suppose that the sales last some time, and that the profit is acquired after the time T_0 : $P(p) = (p - c) D(p) \times (p, T_0)$. If $p = p_m \partial P/\partial p(p_m) = (p_m - c) D(p_m) \partial/\partial p [x(p, T_0)]$. This expression is negative, as it is seen from (1): when p increases, k decreases.

It means that the seller decreases the price to maximize his profit. This is beneficial to consumers (compared to the other choice, perfect information). The explanation is simple: by decreasing the price the seller obtains faster proceeds and wins more money during the time T_0 .

This is confirmed if we consider a discount rate $\delta (0 < \delta < 1)$. It means that at the instant $t = 0$ a flow of proceeds q at the instant t has the value $q \delta^t$. The formula is: $P(p) = (p - c) D(p) F(p)$, with $F(p) = \int_0^{\infty} dx/dt \delta^t dt$. The result is obvious. There is no calculation to make. If p decreases ($p' < p$) the curve $x(p', t)$ is deduced from the curve $x(p, t)$

thanks to an affinity the coefficient of which is $D(p)/D(p') < 1$. The same small part of the proceeds (at a level x, dx) is acquired in advance. Therefore the present value is more.

We conclude that the seller decreases his price when he uses buzz in these conditions: (1) the investment lasts some time and there is no discount rate (2) there is a discount rate and the investment lasts a long time.

Videogames are experience based goods which can be sold only using buzz. Often the sellers provide the first version for free. It is to accelerate the buzz: in (1) the constant k is maximal when $p = 0$. After, the game is known and the seller is permanently in touch with the customers thanks to Internet. The customers become regular. The vendor upgrades the game. The successive versions are sold at some price, which allows profit.

4 CASE 3: THE SELLERS USE BUZZ AND THE CONSUMERS ARE “CURIOUS”

We call these sellers E_1 and E_2 , their costs being c_1 and c_2 . They choose the prices p_1 and p_2 . We consider the “diagram of the utilities”: in a plane $O u_1 u_2$ the utilities of the consumers are represented by a point (u_1, u_2) , u_1 being the utility of the product 1 and u_2 being the utility of the product 2, for this consumer. We suppose $0 \leq u_1 \leq 1$ and $0 \leq u_2 \leq 1$, without any loss of generality. We define $p(u_1, u_2)$: the probability that the utilities (u_1, u_2) of a consumer are such that $u_{10} \leq u_1 \leq u_{10} + du_1$ and $u_{20} \leq u_2 \leq u_{20} + du_2$ is $p(u_{10}, u_{20}) du_1 du_2$.

Then we consider three areas:

A_1 corresponds to the consumers making the purchase of the product 1, if they are informed about the two products: $u_1 - p_1 \geq u_2 - p_2$ and $u_1 - p_1 \geq 0$.

A_2 corresponds to the consumers making the purchase of the product 2, if they are informed about the two products: $u_2 - p_2 \geq u_1 - p_1$ and $u_2 - p_2 \geq 0$.

A_3 corresponds to the consumers making no purchase: $u_1 - p_1 \leq 0$ and $u_2 - p_2 \leq 0$.

Now we call d_1 and d_2 the weights of the two areas A_1 and A_2 : $d_1 = \int_{A_1} p(u_1, u_2) du_1 du_2$, $d_2 = \int_{A_2} p(u_1, u_2) du_1 du_2$ (See Figure 1 where the point M_1 belongs to A_1 , the point M_2 belongs to A_2 and the point M_3 belongs to A_3).

The demands $D_1(p_1, p_2)$ and $D_2(p_1, p_2)$ of the two products are $D_1 = N d_1$ and $D_2 = N d_2$.

We can write the formulas of the profits (in the case of perfect information):

$$P_1(p_1, p_2) = (p_1 - c_1) D_1(p_1, p_2)$$

$$P_2(p_1, p_2) = (p_2 - c_2) D_2(p_1, p_2)$$

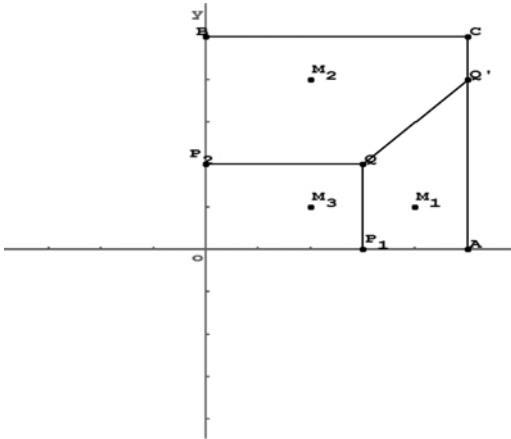


Figure 1. The three areas A_1 , A_2 and A_3 are shown.

The equations giving p_1 and p_2 corresponding to Nash equilibrium are:

$$\partial P_1 / \partial p_1 = 0 \text{ and } \partial P_2 / \partial p_2 = 0.$$

Here we have to explain why Nash equilibrium is interesting when it exists.

When Nash equilibrium exists, is single and stable, it is the only state of the economic sector which is stable: there is no reason for the players to change their choices. But equilibrium only would be too simple [3]. The economic sector is described in a better way by the term “equilibrium of equilibriums” [3]. The two other possible states are price wars and tacit collusion:

Tacit collusion consists in high prices chosen (without any agreement) when the sector is in bad condition. Of course it allows higher profits.

Price war is chosen by one player to end tacit collusion because his profit increases when he decreases his price and gets a big market share. But when he is imitated all the profits in the sector will decrease and soon or later the price war will end [6].

The two states, price war and tacit collusion are unstable, while Nash equilibrium is stable. The sector oscillates between the two unstable states and is often at Nash equilibrium. Therefore, Nash equilibrium is a coarse indicator of the state of the economic sector.

This kind of competition, when the competitors choose their prices, is called Bertrand competition and is well known in game theory [5].

There are a few conditions concerning the diagram of the utilities, D_1 and D_2 , allowing Nash equilibrium which exists, is single and stable [2]. We suppose that these conditions are fulfilled. More, the reaction functions $R_1(p_2)$ and $R_2(p_1)$

have positive slopes. The prices are “strategic complements” meaning that if a player increases his price, the response of the other is to increase his own price.

The definitions of the reaction functions are:

$$p_1 = R_1(p_2), p_1: \text{Max}_{p_1} P_1(p_1, p_2)$$

$$p_2 = R_2(p_1), p_2: \text{Max}_{p_2} P_2(p_1, p_2)$$

($p_1 = R_1(p_2)$ is the best response of the player 1 to p_2 , that is to say he maximizes his profit).

Also, the stable equilibrium should be chosen by short seeing firms (meaning firms anticipating only the consequences of one step).

One checks: $\partial D_1 / \partial p_2 = \partial D_2 / \partial p_1$, therefore the integral $\int -D_1 dp_1 - D_2 dp_2$ can be calculated between two points (u_1, u_2) and (u'_1, u'_2) along any path. It gives the variation of the consumers ‘surplus.

Now we model the buzz. If x ($0 \leq x \leq 1$) is the proportion of informed consumers there are $x(D_1 + D_2)$ agents of diffusion of the buzz. When a consumer is reached by the buzz he will buy the product corresponding to the maximal net utility, no matter the product bought by the consumer having informed him. Or he buys nothing if the two net utilities are negative. Therefore:

$$N dx = k_0 \times (D_1 + D_2) N (1 - x) dt.$$

If $k = k_0(D_1 + D_2)$ we find the same equation than in case 2: $dx = k \times (1 - x) dt$.

After some time, $x \approx 1$ and the two profits are:

$$P_1 = (p_1 - c_1) D_1$$

$$P_2 = (p_2 - c_2) D_2.$$

The outcome of using the buzz, when the consumers are “curious” is the same than when “perfect information” is used.

5 CASE 4: THE SELLERS USE BUZZ AND THE CONSUMERS ARE LAZY

The firms in competition sell their products thanks to buzz, the consumers being “lazy”. When a consumer not yet reached by the buzz meets a purchaser of the product 1, he buys it, or does not buy it. His carelessness has two aspects: (1) if he buys the product 1 he could have preferred to buy the product 2 and (2) if he does not buy the product 1, he could have bought the product 2, which he does not because he is not informed about the product 2. He pays attention to the buzz a single time. He will no more be interested in the buzz. Therefore the equations describing the diffusion of the buzz are:

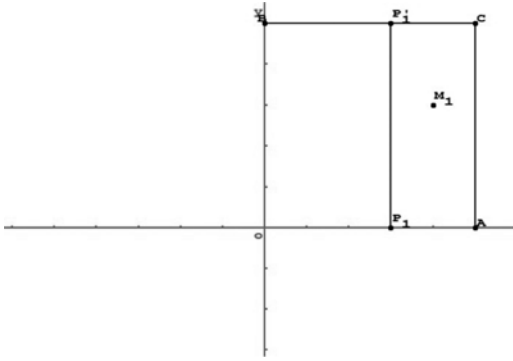


Figure 2. The area $u_1 - p_1 \geq 0$, where is the point M_1 , has a weight W_1 .

$$N dx = k_0 [N \times W_1] N (1 - x - y) dt$$

$N dy = k_0 [N \times W_2] N (1 - x - y) dt$. The proportions of the consumers who are reached by the buzz about the two products are x and y . W_1 and W_2 are the weights of the areas in the diagram of the utilities corresponding to $u_1 - p_1 \geq 0$ and $u_2 - p_2 \geq 0$.

This is shown in the Figure n° 2.

These equations are of the kind "competitive Lotka Volterra" and the solution is:

$(x/x_0)^{W_1} = (y/y_0)^{W_2}$, $x = x_0 f(t)^{W_1}$, $y = y_0 f(t)^{W_2}$, $f(t)$ being the solution of the equation:

$$f'(t) = k f(t) [1 - x_0 f(t)^{W_1} - y_0 f(t)^{W_2}].$$

After a time long enough $f(t) = 1$, given by: $1 - x_0 I^{W_1} - y_0 I^{W_2} = 0$.

The values x_0 and y_0 are the values of x and y at the start ("sowing"). There are a small number of consumers reached by the buzz at the start. To simplify we shall suppose $x_0 = y_0 = k$, k being small. If $k < 1/2$, $K = 1/k > 2$ and $1 > 1$, $I^{W_1} > 1$, $I^{W_2} > 1$, since: $I^{W_1} + I^{W_2} = K$.

The formulas giving the profits are:

$$\begin{aligned} P_1 &= N (p_1 - c_1) D_1(p_1, 1) x_0 I^{D_1(p_1, 1)} \\ P_2 &= N (p_2 - c_2) D_2(1, p_2) y_0 I^{D_2(1, p_2)}. \end{aligned} \quad (2)$$

Here $D_1(p_1, 1)$ is the same than W_1 and $D_2(1, p_2)$ is the same than W_2 .

The value of I is given by $I^{W_1} + I^{W_2} = K$ (if $x_0 = y_0 = k$).

Now we examine the topic of Nash equilibrium: $\partial P_1 / \partial p_1 = 0$, $\partial P_2 / \partial p_2 = 0$.

The function $\partial P_1 / \partial p_1$ is given by (we suppose $N = 1$ and we neglect k):

$$\partial P_1 / \partial p_1 = \partial(p_1 - c_1) w_1 / \partial p_1 I^{W_1} + (p_1 - c_1) w_1 \partial I^{W_1} / \partial p_1. \quad (3)$$

In the general case the calculations are very complex. We can demonstrate the existence of Nash equilibrium (in pure strategies). The reaction function $R_1(p_2)$ is defined (see the appendix for details). But we must suppose that the Nash equilibrium is single.

But the symmetrical case is very interesting because the calculations are easier. It is when the diagram of the utilities is symmetrical ($p(u_1, u_2) = p(u_2, u_1)$) and the costs are the same ($c_1 = c_2 = c$).

One can demonstrate that there is a single Nash equilibrium (see details in the appendix).

Since a single Nash equilibrium exists in the symmetrical case, it allows the comparison between the outcomes of perfect information (used as a selling method) and buzz. We compare the characteristics of the two equilibriums.

First, the prices when buzz is chosen are not very high. They are between the cost c and the price p_m chosen by a firm which is the single in the market (the details of the demonstration are in the appendix).

But the disadvantages of buzz appear if we compare the prices (of the two equilibriums):

If the prices are higher, it is sure that the consumers' surplus decreases.

There are several reasons:

1. Even if we do not take into account x and y (which are equal to $1/2$ in the symmetrical case) the consumers' surplus is less. It is obvious if we use the integral $\int -D_1 dp_1 - D_2 dp_2$ between the two points (p_1, p_2) and (p'_1, p'_2) , $p'_1 > p_1$, $p'_2 > p_2$. Or we can reason on a single increase of p_1 or p_2 (it is obvious that the consumers' surplus decreases).
2. Some consumers lack. The number of lacking consumers is $N (D_1 + D_2 - w_{1x-w_2y})$. No customer lacks in the area $u_1 - p_1 > 0$ and $u_2 - p_2 > 0$. But customers lack in the areas $u_1 - p_1 > 0$, $u_2 - p_2 < 0$ and $u_1 - p_1 < 0$, $u_2 - p_2 > 0$.
3. Among the consumers represented by points in the area $u_1 - p_1 > 0$ and $u_2 - p_2 > 0$ all make a purchase but some should have preferred the product they have not bought. For instance a consumer who has been reached by the buzz about the product 1 has bought it, but he would have preferred to buy the product 2, should he have been informed about this product (it corresponds to $u_2 - p_2 > u_1 - p_1$ and $u_1 - p_1 > 0$).

If the prices decrease, the profits of the sellers decrease.

The point representing the equilibrium moves on the bisector towards the origin of the axis, O .

If we do not take into account x and y , the profits decrease. It is easy to prove. We consider $P_1(p, p)$ with $p < p_N$, p_N being the equilibrium price ($p_1 = p_2 = p_N$). The derived function $d/dp P_1(p, p)$ is positive because it is the sum of two positive terms: $d/dp P_1 = \partial/\partial p_1 P_1 + \partial/\partial p_2 P_1$. More, the values of x and y being $1/2$, the profits (at the Nash equilibrium corresponding to buzz) are less than when the prices are the same, perfect information being chosen.

More accurately, one consumer on two lacks to the firm E_1 in the area $u_1 - p_1 > 0$, $u_2 - p_2 < 0$. And one consumer on two lacks to the firm E_2 in the area $u_2 - p_2 > 0$, $u_1 - p_1 < 0$.

Our conclusion is that choosing the buzz makes always one party (sellers or consumers) lose. Either the consumers' surplus decreases (if prices increase). Either the profits of the sellers decrease (if prices decrease). And possibly, the two parties are losing.

6 CONCLUSION

We shall deal with three topics: (1) we resume the comparison between the outcomes of using "perfect information" or buzz. This is the particular case of a single pricing. There is also the case of successive pricings (2) we treat the case of successive pricings (3) we end by dealing with a particular example, videogames. Also, we deal with the topic of future work.

6.1 Resuming the comparison between "perfect information" and buzz

The symmetrical case allows the comparison between the outcomes of using perfect information or buzz. At least one party (sellers or consumers) is losing. If the prices of Nash equilibrium when buzz is used, are higher, the consumers' surplus is less. If these prices are lower, the sellers' profits are less. But this is the particular case of a single pricing. The outcome of the buzz depends on the prices. They can be chosen a single time. Then the consumers are definitively informed either about the product 1 either about the product 2. If there are successive pricings, our model does not describe the consequences.

Does the stability of Nash equilibrium matter when there is a single pricing? Not really. The stability of the equilibrium matters when short sighted firms can grope. Our model supposes sellers computing data then choosing the equilibrium prices, because it is the only choice that they will not regret, given the other's choice. In other words, it is the best, given the other's choice.

6.2 The case of several pricings

If there are successive pricings we can suppose that the first pricing is (p_{1N}, p_{2N}) which corresponds to Nash equilibrium. Then the values of x and y are definitively $x_N = x(p_{1N}, p_{2N})$ and $y_N = y(p_{1N}, p_{2N})$. The formulas for the profits (it concerns the second pricing, the third pricing etc.) are:

$$P_1 = N(p_1 - c_1) w_1 x_N$$

$$P_2 = N(p_2 - c_2) w_2 y_N$$

For instance the good has to be bought again because of obsolescence.

The pricing should be (p_{1m}, p_{2m}) . These prices correspond to the price chosen when one of the two sellers is the single in the market (that is to say when the other chooses the price 1).

Therefore:

$$P_1 = N(p_{1m} - c_1) w_1 x_N$$

$$P_2 = N(p_{2m} - c_2) w_2 y_N$$

In the symmetrical case only one is sure that there is a single Nash equilibrium (during the first phase, when the buzz occurs), and the formulas are: $P_1 = 1/2 N(p_m - c) w$ and $P_2 = 1/2 N(p_m - c) w$.

It is not Nash equilibrium. The profit of one seller does not depend on the choice of the other. Each seller maximizes his profit as if he was alone in the market. In the formulas (2) the profit of one seller depends on the choice of the other because l depends on the two prices ($l^w + l^{w^2} = K$ means that l is a function of p_1 and p_2).

If we compare the outcomes of "perfect information" and buzz when there are successive pricings, the conclusion is that the consumers' surplus is less when buzz is chosen: (1) the prices are higher, since the equilibrium prices of the Bertrand competition are lower than p_{1m} and p_{2m} and (2) one has to take into account x_N and y_N .

6.3 The example of videogames

Videogames are experience based goods which should be sold only using buzz. When the first version is free in any case: $x = 1/2$ and $y = 1/2$. When the prices are 0, no consumer lacks, since there is no areas $u_1 - p_1 \geq 0$, $u_2 - p_2 \leq 0$ or $u_1 - p_1 \leq 0$, $u_2 - p_2 \geq 0$. But it is another story when there is the second pricing, the third pricing etc. The seller knows his customers, is permanently in touch with them, upgrades his product and chooses prices allowing profit. The sellers should choose p_{1m} and p_{2m} :

$$P_1 = 1/2 N(p_{1m} - c_1) w_1$$

$$P_2 = 1/2 N(p_{2m} - c_2) w_2$$

Customers lack: one on two in the area $u_1 - p_{1m} \geq 0, u_2 - p_{2m} \leq 0$ (concerning the seller 1) and in the area $u_1 - p_{1m} \leq 0, u_2 - p_{2m} \geq 0$ (concerning the seller 2).

But this lack of consumers (compared to the consumers who would have bought the products, should perfect information have been used) is inevitable. A videogame, being an experience based good, can be sold only thanks to buzz.

Therefore, if a seller spends very much money when he creates the game, it is not to acquire more customers. In any case he will have $N/2$ customers. But these customers will have higher utilities. Later, when the seller chooses p_{1m} , this price can be higher. It allows more profit.

And why to deliver the first version freely? It is not to acquire more customers: later, when the price p_{1m} will be chosen, all the customers having downloaded the game, whose utility is less than p_{1m} ($u_1 - p_{1m} < 0$) will lack. But the freeness allows a faster buzz.

The sellers avoid any competition. But it is not deliberate. Again, videogames being experience based goods can be sold only using buzz. In the first phase delivering the game is free, because it makes the second phase (when the pricing allows profit) occur sooner. In this phase a seller can upgrade the game and sell it at higher price, but he keeps his customers. He is in touch with them and can inform them about the new versions of the game. He cannot gain customers from the other seller, since he is not in touch with them.

6.4 Future work

Future work could concern the method and advertising in the case of videogames:

The method. What is interesting in the method we used is the “diagram of the utilities”. We already used it in the Chapter 6 of our book “Game theory and the stakes in the telecommunications industry” [2]. The diagram of the utilities could be a good method to deal with other topics, if Bertrand competition is concerned.

Advertising in the case of videogames. Here empirical studies could be interesting. Indeed, to advertise videogames is not so frequent. In France, for instance, to advertise videogames on TV is forbidden. In Germany, to advertise a violent game which is called “Barbarian: the ultimate warrior” is forbidden. The topic is awkward: when a game is censored, or to advertise it is forbidden, there is a controversy and ... it makes the game renowned. But there is some advertising (example are Dark Age of Camelot, Call of Duty, Grand Theft Auto etc.), mainly on Internet. Interesting questions are: (1) What is the gain when a seller advertises his

game? (2) What is the loss when a game cannot be advertised because it is forbidden?

7 APPENDIX

We start from $l^\alpha + l^\beta = K$ (we have replaced w_1 by α and w_2 by β). When p_1 decreases (from 1 to c_1), α increases. The quantity l decreases: $\partial l / \partial \alpha = -l^\alpha \log l / \alpha l^{\alpha-1} + \beta l^{\beta-1} < 0$ (since $l > 1$). But the quantity l^α increases: $\partial l^\alpha / \partial \alpha = l^{\alpha \beta} \log l / \alpha l^{\alpha} + \beta l^\beta > 0$. The number of customers of E_1 , $Nk\alpha l^\alpha$ increases and the number of customers of E_2 , $Nk\beta l^\beta$ decreases.

Now we consider (3). When $p_1 = c_1$, $\partial P_1 / \partial p_1$ is positive and when $p_1 = p_{1m}$ (the single maximum of $P_1 = (p_1 - c_1) w_1$ is when $p = p_{1m}$), it is negative, because $\partial l^\alpha / \partial p_1 < 0$.

When p_2 has any value, $(p_1 - c_1) w_1$ has a single maximum, $\partial^2 (p_1 - c_1) w_1 / \partial p_1^2$ being negative (it is a sufficient condition). It is Bertrand competition when $p_2 = 1$ and the profit is a concave function. Therefore there is a single zero or several zeros when $c_1 < p_1 < p_{1m}$. If $p_1 > p_{1m}$, $\partial P_1 / \partial p_1$ is negative and a zero is impossible.

To demonstrate that there is a single zero we write the equation $\partial P_1 / \partial p_1 = 0$:

$$l^{w_1} [\partial (p_1 - c_1) w_1 / \partial p_1 + (p_1 - c_1) w_1 / l^{w_1} \partial l^{w_1} / \partial w_1 \partial w_1 / \partial p_1] = 0.$$

(Here we neglect the constants N and k .)

It is enough to demonstrate that the member at left is decreasing (when p_1 varies from c_1 to p_{1m}):

1. $\partial (p_1 - c_1) w_1 / \partial p_1$ decreases
2. $\partial w_1 / \partial p_1$ is negative and its absolute value increases ($\partial^2 w_1 / \partial p_1^2 < 0$ because it is a sufficient condition for $\partial^2 (p_1 - c_1) w_1 / \partial p_1^2 < 0$).
3. the other factors $(p_1 - c_1) w_1$ and $F = 1/l^{w_1} \partial l^{w_1} / \partial w_1$ are positive and increasing. The product of three factors $[(p_1 - c_1) w_1] [F] [\partial w_1 / \partial p_1]$ is negative, and its absolute value increases. Therefore it is decreasing. It is enough to demonstrate that $\partial F / \partial p_1$ is positive, or $\partial F / \partial \alpha$ is negative, since $\partial F / \partial p_1 = \partial F / \partial \alpha \partial \alpha / \partial p_1$, and $\partial \alpha / \partial p_1 < 0$.

Replacing l^α and l^β by X and Y for simplicity:

$$F = 1/X [XY \log Y / \alpha X + \beta Y]$$

$$F = Y \log Y / \alpha K + (\beta - \alpha) Y.$$

The derived in α (β is supposed fixed and Y is function of α) is:

$$\partial F (\alpha, Y) / \partial \alpha + \partial F (\alpha, Y) / \partial Y \partial Y / \partial \alpha.$$

A simple calculation shows $\partial F / \partial \alpha < 0$ and $\partial F / \partial Y > 0$. Also $\partial Y / \partial \alpha = -\partial X / \partial \alpha < 0$.

Since there is a single zero of $\partial P_1/\partial p_1$ the reaction function $R_1(p_2)$ exists.

But in the general case, if we want to know if there is a single intersection point of R_1 and R_2 , the calculations are too complex. We can make the hypothesis that there is a single Nash equilibrium.

But there is a particular case, the symmetrical case, which allows demonstrating there is a single Nash equilibrium. It is defined: the diagram of the utilities is symmetrical with respect to the bisector, and the costs c_1 and c_2 are the same. Any Nash equilibrium is on the bisector.

The value of $\partial P_1/\partial p_1$ on the bisector (multiplied by a constant K) is:

$$K \cdot \frac{\partial P_1}{\partial p_1} = \frac{\partial(p_1 - c_1)}{\partial \alpha} \frac{\alpha}{\partial p_1} + (p - c) \frac{1}{2} \log K/2$$

On the bisector, $p_1 = p_2$ and $\alpha = \beta$. One supposes p increasing from c_1 to p_{1m} . There is a single point corresponding to $\partial P_1/\partial p_1 = 0$. The function is positive when $p = c_1$, negative when $p = p_{1m}$, and decreasing.

We have demonstrated that in the symmetrical case there is a single Nash equilibrium. We do not know if it is stable but this does not matter (see in the conclusion the paragraph 6.1).

REFERENCES

- [1] Kahneman D (2011) Thinking fast and slow. Farrar, Strauss and Giroux.
- [2] Lefebvre O (2014) Game theory and the stakes in the telecommunications industry. Lambert Academic Publishing.
- [3] Morin E (1981) The nature of Nature. Editions du Seuil.
- [4] Moulin R (1967) The market of painting in France. Editions de minuit.
- [5] The Palgrave Dictionary of Economics (2008) Article "Bertrand competition" Authors Baye R and Kovenock D Macmillan Publishers.
- [6] Tirole J (1988) The theory of industrial organization. MIT Press.

Verification of merge sorting technique using formal methods

Nida Mahmoud Zaitoun & Musbah J. Aqel

Faculty of Information Technology, Applied Science University, Amman, Jordan

ABSTRACT: In this paper, the merge sorting technique is represented as predicate logic which is one of the popular approaches of formal methods. Moreover, the representation of the merge sorting techniques and the required predefined functions (predicates) was verified using Hoare logic.

The results of the verification of this representation proved that the representation method is true. The representation and verification will be very useful for software engineering which will enhance the software.

1 INTRODUCTION

1.1 Verification

One of the most important aspects of software development process is validation and verification as shown in Figure 1. Validation is the process to ensure that the software meets the user's requirements. While the Verification is the process to ensure reliable software by testing and bugs removing [1, 3, 4, 14].

The software engineer Boehm described the difference as follows [1, 3]:

- Verification: "Are we building the product right?"
- Validation: "Are we building the right product?"

Acceptance Testing is an example of validation where the software is apparently complete and demonstrated to its clients and accepted by them. The Unit Testing and System Testing are examples of verifications. In the Unit Testing each module of the software is tested in isolation. The modules will then be linked to each other to form a system. This complete system will be tested during the system testing phase [1, 3, 4, 14].

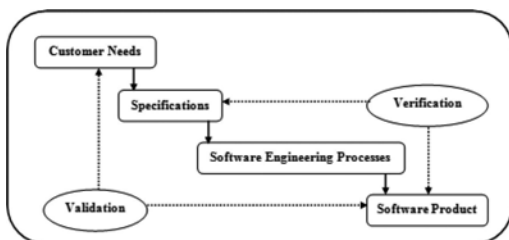


Figure 1. Software verification and validation.

The Verification of a program is to prove the program's correctness with respect to some specification [1, 3, 14, 19, 28]. While the Derivation is constructing only correct programs from the specification [14, 20].

1.2 Verification Tasks

The verification tasks include the following:

A. Errors

A software system has an overall specification derived from the requirement analysis, and each software component has an individual specification derived from architectural design [1, 3].

The specification of a component can be ambiguous, incomplete or faulty. Ambiguous and incomplete may lead to another problem during programming which is misunderstand the component specification. Faulty may be due to kinds of errors. That said, it is clear that specifications are the common source of faults [1, 3, 14, 19, 26].

B. Testing

"Testing can demonstrate the presence of bugs, but it can never prove their absence" is the famous statement of E.W. Dijkstra [23, 24].

Testing a program is not only looking at the program, looking at a program is called inspection or walkthrough. Testing methods are different from each other in many points such as inputs and limitations [1, 3, 14, 16, 19, 27].

Any program is likely to contain bugs. This fact forces the software developer to use different testing strategies such as Black Box Testing and White Box Testing, Unit Testing and System Testing, or other testing strategies as stepping through code, testing the test data, team technique and beta testing [1, 3, 4, 14].

C. Formal Software Verification

Formal verification of a computing system implies a mathematical proof of correctness about computer programs showing that the system satisfies its specification. Accordingly, some mathematical structure to model the system must be used [4, 7, 8, 14, 16, 19, 20, 21, 23, 24, 25, 26, 28, 29].

Floyd-Hoare Triple presented how to formalize program's correctness by two predicates: firstly, the precondition of the program which represents the pre states to start the program, and secondly, the post condition of the program in which the program should terminate. All other predicates (assertions) are called Intermediate Predicates (assertions) which serves as steps of reason between precondition and post condition. The triple precondition, program, and post condition, is called a Floyd-Hoare Triple or Hoare Rules [4, 7, 12, 14, 17, 19, 20, 21, 23, 24, 25, 29]

2 FORMAL METHODS

Applying mathematics in software development is known as "Formal Methods" which means connecting program code and its logic specifications. Second meaning of formal methods is model checking that is used to reason the correctness of hardware circuits and software systems. A third meaning of formal methods is to describe software systems using specification languages, such as Z or VDM. Formal languages' semantics is expressed in logic [1, 3, 5, 6, 9, 13, 15, 16, 18].

2.1 Logics

Logics are the languages of mathematics used to formally capture the concepts about which one wishes to reason [10, 11, 13, 18] and logical expressions have a precise meaning by producing a value of true or false for every assignment of values to the variables in an expression [2, 10, 11, 22].

Thus, the base of FM is logics of various kinds and in these logics one defines models, expresses properties about models, and then performs reasoning about the model and its properties. As a result studying formal method implies a great concern with logics, reasoning techniques and tools, and standardized languages [16, 18].

2.2 Reasoning

Logics are used to describe the objects about which one wants to reason. The reasoning itself is an attempt to construct a proof of the desired claim. The term proof is either a set of steps that adheres to the pertaining calculus or it is an exhaustive enumeration of all the possible valuations [4, 14, 18].

i. Program Reasoning Using Hoare's Logic

Running some tests on sample input proves that a given code works correctly and usefully, however it cannot prove that the code works in all possible scenarios because there is an infinite number of possible test data. This is where reasoning about block code comes in. Formally reasoning about code is the process that enables to prove that a given code is working correctly using logical formulas to show what must be true at each point in the code and these logical formulas are called assertions.

Hoare Triple is the formalization form of all assertions [12, 19, 20, 21, 23, 24, 25, 29]. Floyd-Hoare Triple presents how to formalize program's correctness. The triple, precondition, program and post condition, is called a Floyd-Hoare Triple or simply Hoare Triple or Hoare Rules [7, 17, 19, 20, 21, 23, 24, 25, 29]. It is useful to make advantage of the power of the Floyd/Hoare logic of imperative programs [8, 20].

2.3 Standardized languages

Many of the input languages are tools that have been standardized to avoid tool "lock-in" and enables the use of multiple tools to solve (various different parts of) a problem.

3 MERGE SORTING TECHNIQUE USING FORMAL METHODS FOR REPRESENTATION

The idea of this research emerged from the fact that the formal methods provide a formal (exact and unambiguous) notation and they provide verification mechanisms for checking that the codes produced do really implement the requirements. Depending on the previous facts or properties of formal methods, predicate logic is used to produce an exact and unambiguous representation of merge sorting technique, and Hoare logic is used but as a verification mechanism for checking that the predicate logic representations implements the requirements of sorting techniques.

The translation approach was started with defining logical operations, and all along the research, the predefined functions (predicates) were defined when it was needed. The merge sorting technique was analyzed in order to reengineer and verify it; listing the statements and pseudocode of the technique then the predicate representation, the predicate representation consists of two parts: first part is the definitions and rules, and the second part is the sorting technique representation into predicate logic.

The pseudocode of the merge sorting technique was analyzed in order to convert it from sequence of steps into recursive or inductive definition with

hierarchy structure i.e. to convert the algorithm from a sequence of steps into using some subroutines, and instead of using loops in implementation of the algorithms, we redefined the sorting technique using recursive definition.

The predicate logic representation starts with the identification of the three clauses of recursive or inductive definition such as: first clause is the Basis which defines the set of the 'basis' elements, the second clause is the Inductive clause or Induction which specifies the ways in which new elements can be produced from the 'basis', and third clause is the External clause which asserts that the basis and induction must be applied by the element before being a member of the set. Later on we determined the input and the output of the algorithm before representing the algorithm using predicate logic.

The final stage of the research is the verification using Hoare logic, using assertions in form of preconditions and post conditions as Hoare triple form which is written as: {Preconditions} Statement {Post conditions} in simple form as {P} S {Q}. Predefined functions in addition to sorting techniques that have been verified using Hoare logic.

3.1 Statements of merge sort (Divide-and-Conquer)

1. Divide set into two halves
2. Recursively sort each half.
3. Merge two halves to make sorted whole, the steps of merging the two halves are:
 - o Keep track of smallest element in each sorted half.
 - o Insert smallest of two elements into auxiliary set.
 - o Repeat until done.

3.2 Pseudocode of merge sorting algorithm

- i. Pseudocode of the Merge sorting algorithm using some required predefined functions:


```

/* Split in half
m = n/2
if Left_Part = [] then {
  Left_Part = Set_S [1.. m]
/* Insertion sort for sorting two parts of S
  Insertion_Sort (Left_Part)
  Insertion_Sort (Set_S [m+1.. n])
/* Merge sorted sub-sets
  i = 1, j = m+1, k = 1}
while i <= m and j <= n,
Sk++ = (Si < Left_Partj)? Sj++: Left_Parti++
→ invariant: S[1..k] in final position
while i <= m,
  Sk++ = Left_Parti++
→ Invariant: S [1..i] is sorted.
```

- ii. Pseudocode of recursive definition of the Merge sort algorithm:

```

m = n/2
if Left_Part = [] then {
  Left_Part = Set_S [1.. m]
/* Insertion sort for sorting two parts of S
  Insertion_Sort (Left_Part)
  Insertion_Sort (Set_S [m+1.. n])
/* Merge sorted sub-sets
  i = 1, j = m+1, k = 1}
if ( i ≤ m) and (j ≤ n) and (Sj < Left_Parti) then
  Sk++ = Sj++
  Merge (S, Left_Part, i, k)
→ invariant: S[1..k] in final position
if ( i ≤ m) then
  Sk++ = Left_Parti++
  Merge (S, Left_Part, i, k)
→ Invariant: S [1..i] is sorted.
```

3.3 Predicate logic representation of merge sorting technique

3.3.1 Definitions and rules

Cardinality (S) ↔ |S|
 Equal (X, Y) ↔ X = Y
 GE (X, Y) ↔ X ≥ Y
 Is (X, Y) ↔ Sets X = Y
 LE (X, Y) ↔ X ≤ Y
 Less (X, Y) ↔ X < Y
 Post_Element (X_i) ↔ X_{i+1}
 Insertion_Sort
 Left_Split

3.3.2 Representation of merge sorting algorithm

Basis Clause:

Set_S: is a set of unsorted n elements that needs to be sorted using Merge sorting technique based on predicate logic and needs to be verified by Hoare logic.

Left_Part: is a temporary set using to copy left half of the Set_S to be used in merge stage such that Left_Part = Set_S [1..n/2]

Initial_i = 1, Initial_k = Initial_i,
 Merge_Sort (Set_S, Left_Part, Initial_i, Initial_k)

- o Basis Clause using pseudocode:


```

Merge(S, L, i, k)
m = n/2
if L = [] then {
  L = S [1..m]
  Insertion_Sort (L)
  Insertion_Sort (S [m+1..n])}
```
- o Basis Clause using predicate logic:


```

Merge (S, L, i, k)
Is (m, Cardinality (S)/2)
Equal (L, []) → Left_Split (S, L)
Equal (L, []) → Insertion_Sort (S, Cardinality (L) +2, Cardinality (S), i, 1)
```


Equal (L, []) \rightarrow Insertion_Sort (L, 2, Cardinality (L), 2, 1)

Inductive Clause:

- o Inductive Clause using pseudocode:
 - while $i \leq m$ and $j \leq n$,
 - $S_{k++} = (S_j < L_i)? S_{j++} : L_{i++}$
 - while $i \leq m$,
 - $S_{k++} = L_{i++}$
- o Inductive Clause using predicate logic:
 - LE (i, Cardinality (L)) \wedge LE (j, Cardinality (S))
 \wedge Less (S_j, L_i) \rightarrow Equal (Post_Element (S_k),
 Post_Element (S_j))
 - LE (i, Cardinality (L)) \wedge LE (j, Cardinality (S))
 \wedge Less (S_j, L_i) \rightarrow Merge (S, L, i, k)
 - LE (i, Cardinality (L)) \wedge LE (j, Cardinality (S))
 \wedge GE (S_j, L_i) \rightarrow Equal (Post_Element (S_k),
 Post_Element (L_j))
 - LE (i, Cardinality (L)) \wedge LE (j, Cardinality (S))
 \wedge GE (S_j, L_i) \rightarrow Merge (S, L, i, k)
 - LE (i, Cardinality (L)) \rightarrow Equal (Post_Element (S_k),
 Post_Element (L_i))
 - LE (i, Cardinality (L)) \rightarrow Merge (S, L, i, k)

External Clause

S [1.. i] is sorted $\leftrightarrow \forall S_i (i \geq 1 \wedge i \leq |S|) \in S \rightarrow S_i \leq S_{i+1}$

Algorithm Merge_Sort (Set_S, Left_Part, Initial_i, Initial_k)

Input: The Set_S is a set of unsorted n elements, Left_Part: is a temporary set using to copy left half of the Set_S to be used in merge stage such that Left_Part = Set_S [1..n/2]

Output: Sorted Set_S such as: $\forall S_i (i \geq 1 \wedge i \leq |S|) \in \text{Set_S}, S_i \leq S_{i+1}$

Algorithm:
 Merge_Sort (S, L, I, K)
 /* Split in half
 Is (m, Cardinality (S)/2)
 Equal (L, []) \rightarrow Left_Split (S, L)
 /* Insertion sort for sorting two parts of S
 Equal (L, []) \rightarrow Insertion_Sort (S, Cardinality (L) +2, Cardinality (S), I, 1)
 Equal (L, []) \rightarrow Insertion_Sort (L, 2, Cardinality (L), 2, 1)
 /* Merge sorted sub-sets using temp set
 Equal (L, []) \rightarrow Is (I, 1)
 Equal (L, []) \rightarrow Is (J, Cardinality (L) +1)
 Equal (L, []) \rightarrow Is (K, 1)
 LE (I, Cardinality (L)) \wedge LE (J, Cardinality (S))
 \wedge Less (S_j, L_i) \rightarrow Is (Post_Element (S_k),
 Post_Element (S_j))
 LE (I, Cardinality (L)) \wedge LE (J, Cardinality (S))
 \wedge Less (S_j, L_i) \rightarrow Merge_Sort (S, L, I, K)
 LE (I, Cardinality (L)) \wedge LE (J, Cardinality (S))
 \wedge GE (S_j, L_i) \rightarrow Is (Post_Element (S_k),
 Post_Element (L_j))
 LE (I, Cardinality (L)) \wedge LE (J, Cardinality (S))
 \wedge GE (S_j, L_i) \rightarrow Merge_Sort (S, L, I, K)

\rightarrow invariant: S[1..k] in final position
 LE (I, Cardinality (L)) \rightarrow Is (Post_Element (S_k),
 Post_Element (L_i))
 LE (I, Cardinality (L)) \rightarrow Merge_Sort (S, L, I, K)
 LE (I, Cardinality (L)) \rightarrow Is (Post_Element (S_k),
 Post_Element (L_i))
 LE (I, Cardinality (L)) \rightarrow Merge_Sort (S, L, I, K)
 \rightarrow invariant: S [1..i] is in final position (sorted).

3.4 Representation of required functions

3.4.1 Swap

- a. Statements of the Swap algorithm
 Swapping the values of two variable means to exchange their values
- b. Algorithm of the Swap algorithm
 T = X
 X = Y
 Y = T

c. Predicate logic representation

- 1. Definitions and Rules
 Is (X, Y) \leftrightarrow Lets X = Y
- 2. Representation of Swap algorithm

Algorithm Swap (X, Y)
Input: (X = X_{pre}, Y = Y_{pre})
Output: (X = Y_{pre}, Y = X_{pre})
Algorithm:
 Swap (X, Y)
 Is (T, X)
 Is (X, Y)
 Is (Y, T)

3.4.2 Left Split

a. Statements of Left Split:

- 1. Start with a mean set of cardinality n needed to be split and an empty set to be a container of one half on the mean set.
- 2. Then move one half; say left one from the mean set to the temporary one which is of cardinality m = n/2.

b. Pseudocode of Left Split algorithm:

- Pseudocode of sequential steps of the Left Split algorithm
 /* split in half
 m = n/2
 for i = 1:m
 $L_{i++} = S_{i++}$
- Pseudocode of recursive definition of the Left Split algorithm
 /* split in half
 if (i \leq m) then
 $L_{i++} = S_{i++}$
 Left_Split (S, L, i)

c. Predicate logic representation

- 1. Definitions and Rules
 Cardinality (S) \leftrightarrow |S|
 Is (X, Y) \leftrightarrow Lets X = Y
 LE (X, Y) \leftrightarrow X \leq Y

2. Representation of Left Split algorithm

Basis Clause:

S: is a set of unsorted n elements.

L: is a set using to copy left half of S such that

$L = S [1..|S|/2]$,

$m = |S|/2$,

Left_Split (S, L, i)

Inductive Clause:

- Inductive Clause using pseudocode:

$L_{i++} = S_{i++}$

- Inductive Clause using predicate logic:

Is (Post_Element (L_i), Post_Element (S_i))

External Clause

$L = S [i.. n/2] \leftrightarrow \forall i (i \geq 1 \wedge i \leq |S|/2) L_i \in L \rightarrow S_i \in S$

Algorithm Left_Split (S, L)

Input: The S is a set of n elements; L is an empty set.

Output: Set L contains the left half of S such that $L = S [1..|S|/2]$

Algorithm:

Left_Split (S, L, I)

/* Split the left half of the set S into the set L

LE (I, Cardinality(S)/2) \rightarrow Is (Post_Element (L_i), Post_Element (S_i))

LE (I, Cardinality(S)/2) \rightarrow Left_Split (S, L, I++)

4 MERGE SORTING TECHNIQUE USING FORMAL METHODS FOR VERIFICATION

4.1 Verification of Merge_Sort

{Set_S is a set contains n unsorted elements \wedge Left_Part is a temporary set to copy left half of the Set_S to be used in merge stage $\wedge i = \text{Initial}_i = 1 \wedge \text{Initial}_k = \text{Initial}_i$ }

Merge_Sort (S, L, I, K)

{Local variables S, L, i, k: $S \leftarrow \text{Set}_S \wedge L \leftarrow \text{Left_Part} \wedge i \leftarrow 1 \wedge k \leftarrow i = 1$ }

/* Split in half

{ $S = \text{Set}_S \wedge L = []$ }

Is (m, Cardinality (S)/2)

{ $m = |S|/2$ }

{ $L = []$ }

Left_Split (S, L)

/* Insertion sort for sorting two parts of S

Insertion_Sort (S, Cardinality (L) +2,

Cardinality (S), i, 1) \wedge

Insertion_Sort (L, 2, Cardinality (L), 2, 1) \wedge

/* Merge sorted sub-sets using temp set

Is (I, 1)

Is (I, Cardinality (L) +1)

Is (K, 1)

{ $S = \text{Set}_S \wedge \text{Right part of S is sorted set such as:}$

$\forall S_i (i \geq m + 1 \wedge i \leq n) \in \text{Set}_S, S_i < S_{i-1} \wedge L = S [1.. m] \wedge L$ is a sorted set such as: $\forall L_i (i \geq 1 \wedge i \leq m) \in L, L_i < L_{i-1} \wedge i = 1 \wedge j = m + 1 \wedge k = 1$ }

...

/* ith iteration

{ $i < = |L| \wedge j < = |S| \wedge S_j < L_j$ }

Is (Post_Element (S_i), Post_Element (S_j))

Merge_Sort (S, L, I, K)

{ $i < = |L| \wedge j < = |S| \wedge \neg (S_i < L_j)$ }

Is (Post_Element (S_i), Post_Element (L_j))

Merge_Sort (S, L, I, K)

{ $\neg (i < = |L| \wedge j < = |S|)$ }

...

\rightarrow invariant: S[1..k] in final position

{ $i < = |L|$ }

Is (Post_Element (S_i), Post_Element (L_i))

Merge_Sort (S, L, I, K)

{ $\neg (i < = |L|)$ }

{Set_S is a sorted set such as: $\forall S_i \in \text{Set}_S, S_i < S_{i+1}$ }

\rightarrow invariant: S [1..i] is in final position (sorted).

4.2 Verification of Swap

/* Verification of Swap as a function

{ $X = X_{pre} \wedge Y = Y_{pre}$ }

Swap (X, Y);

{ $X = Y_{pre} \wedge Y = X_{pre}$ }

/* Verification step by step of Swap algorithm

{ $X = X_{pre} \wedge Y = Y_{pre}$ }

Is (T, X);

{ $X = X_{pre} \wedge Y = Y_{pre} \wedge T = X_{pre}$ }

Is (X, Y);

{ $X = Y_{pre} \wedge Y = Y_{pre} \wedge T = X_{pre}$ }

Is (Y, T);

{ $X = Y_{pre} \wedge Y = X_{pre} \wedge T = X_{pre}$ }..

4.3 Verification of Left_Split

{Set_S is a set contains n elements \wedge L is an empty set}

Left_Split (S, L, I)

{Local variables S and L such as $S \leftarrow \text{Set}_S \wedge L \leftarrow |S|/2$ }

/* Split the left half of the set S into the set L

...

/* ith iteration

{ $i \leq |S|/2$ }

Is (Post_Element (L_i), Post_Element (S_i))

Left_Split (S, L, I++)

{ $\neg (i \leq |S|/2) \wedge L = S [1.. |S|/2]$ }

...

\rightarrow Invariant: $L = S [i.. |S|/2]$

5 RESULTS

The formal methods provide an exact and unambiguous notation, and they also provide verification mechanisms. As such, Predicate logic is used to produce an exact and unambiguous representation of merge sorting technique, and Hoare logic

is used to verify that the Predicate logic representations carry out the requirements of sorting techniques.

Start with the merge sort which is a divide-and-conquer based sort; listing the statements and pseudocode of the merge sort then determining the definitions and the rules that would be needed in representing the merge sort into predicate logic. The pseudocode of the merge sort was designed using recursive or induction and defining swap and left_split subroutines (predicates). After that, the basis, inductive, and external clauses were identified. The input and the output of the algorithm are determined before the merge sort is represented into predicate logic.

The verification stage is applied on all the merge sorting technique and the predefined functions (predicates) using Hoare logic, Assertions in form of preconditions and post conditions are used to verify the correctness of the predicate logic representation.

The output of the verification has shown that the representation is true.

6 CONCLUSION

The merge sorting technique and required predefined functions (predicates) are subjected to a treatment in order to be represented into predicate logic after that the verification stage is applied on the predicate logic representation of the selected sorting techniques and the predefined functions (predicates) using Hoare logic.

The output of the verification has shown that the representation is true; since the output of the verification stage is proved to be true for the predicate logic representation for the input which is the merge sorting technique and required predefined functions (predicates). This will enhance the software analysis and design and should deliver the required functionality and performance to the user and should be maintainable, dependable and usable.

7 RECOMMENDATIONS

The results of this research i.e. the representation of sorting techniques using predicate logic and verification using Hoare logic are worthy of being used in software engineering applications to design good software which would deliver the required functionality and performance to the user and should be maintainable, dependable, and usable. The formal methods field is a very active research area with a wide variety of methods and mathematical models and sorting techniques field

which useful in all fields of computer science and software engineering.

REFERENCES

- [1] Bell, D. (2000). *Software Engineering A Programming Approach*. 3rd Ed. Addison-Wesley, Pearson Education Ltd.
- [2] Mangalgi, S.R. (1988). *Modern Algebra & Trigonometry*. 2nd Ed. Vidya Prakashan, Nagpur.
- [3] Sommerville, I. (2009). *Software Engineering*. 9th Ed. Pearson.
- [4] Adrion, W. R., Branstad, M. A., & Cherniavsky, J. C. (1982). Validation, verification, and testing of computer software. *ACM Computing Surveys (CSUR)*, 14(2), 159–192.
- [5] Batra, M. (2013). Formal Methods: Benefits, Challenges and Future Direction. *Journal of Global Research in Computer Science*, 4(5), 21–25.
- [6] Bernhard, B. (2006). *Formal Verification of Software*. [Online]. Available: <http://searches.globososo.com/search/web?type=sc&channel=ild&q=Formal%20Verication%20of%20Software%20Bernhard%20Beckert>.
- [7] Bjørner, N., Gurfinkel, A., McMillan, K., & Rybalchenko, A. Horn Clause Solvers for Program Verification.
- [8] Bornat, R. (2000, January). Proving pointer programs in Hoare logic. In *Mathematics of program construction* (pp. 102–126). Springer Berlin Heidelberg.
- [9] Burgess, C. J. (1995). The role of formal methods in software engineering education and industry. *University of Bristol, UK*.
- [10] Critical thinking web. [Online]. Available: <http://philosophy.hku.hk/think/>.
- [11] Dramnesc, I., & Jelelean, T. (2010). Proof based synthesis of sorting algorithms. In *RISC Report Series*. University of Linz Austria. [Online].
- [12] Ferreira, M. A., & Oliveira, J. N. (2009). An integrated formal methods tool-chain and its application to verifying a file system model. In *Formal Methods: Foundations and Applications* (pp. 153–169). Springer Berlin Heidelberg.
- [13] Fisher, M. *Temporal Logic [Introducing Formal Methods]*. [Online]. Available: www.csc.liv.ac.uk/~michael/TLBook/tl1-4up.pdf.
- [14] Gaudel, M. C. (2005). Formal methods and testing: Hypotheses, and correctness approximations. In *FM 2005: Formal Methods* (pp. 2–8). Springer Berlin Heidelberg.
- [15] Hall, A. (2007). Realising the Benefits of Formal Methods. *J. UCS*, 13(5), 669–678.
- [16] Heitmeyer, C. (1998, January). On the need for practical formal methods. In *Formal Techniques in Real-Time and Fault-Tolerant Systems* (pp. 18–26). Springer Berlin Heidelberg.
- [17] Hoare, C. A. R. (1969). An axiomatic basis for computer programming. *Communications of the ACM*, 12(10), 576–580.
- [18] Janota, M., Kiriya, J., & Botterweck, G. (2008). Formal methods in software product lines: concepts, survey, and guidelines. *Lero, University of Limerick, Tech. Rep. TR-SPL-2008-02*.

- [19] Janota, M. (2005). *Automated Theorem Proving and Program Verification*: Master Thesis. Charles University, Prague: Czechoslovakia.
- [20] *Limitations of Propositional Logic*. [Online]. Available: <https://classes.soe.ucsc.edu/cmcs140/Winter11/lectures/Feb7-FOL.pdf>.
- [21] Oliver, I. (2007, December). Experiences of Formal Methods in Conventional Software and Systems Design. In *BCS FACS Xmas Workshop: Formal Methods in Industry (December 2007)*.
- [22] Pandey, S. K., & Batra, M. (2013). Formal Methods in Requirements Phase of SDLC. *International Journal of Computer Applications*, 70(13), 7–14.
- [23] Parnas, D. L. (1993). Predicate logic for software engineering. *Software Engineering, IEEE Transactions on*, 19(9), 856–862.
- [24] *Predicate Logic*. [Online]. Available: <https://www.cs.hmc.edu/~keller/cs60book/10%20Predicate%20Logic.pdf>.
- [25] *Predicate Logic—Symbols, Syntax, Semantics, Translation*. [Online]. Available: http://www.davidagler.com/teaching/logic/handouts/Handout6_PredicateSymbolsSyntaxSemanticsTranslation.pdf.
- [26] *Requirements and Formal Methods*. [Online]. Available: <http://se.inf.ethz.ch/old/teaching/ws2005/0273/slides/formalMethods.pdf>
- [27] Rushby, J. (1989). Formal methods and critical systems in the real world. *Formal Methods for Trustworthy Computer Systems (FM89)*, 121–125.
- [28] School of Computing Edinburgh Napier University. *Predicate logic SET07106 Mathematics for Software Engineering*. [Online]. Available: <http://www.upriss.org.uk/math/mlec10.pdf>.
- [29] Turner, H. (2008). Nonmonotonic causal logic. *Foundations of Artificial Intelligence*, 3, 759–776.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Barriers surrounding e-government implementation: A case study of Government to Business (G2B) system

Hasan Hashim, Angela Lin & Jonathan Foster

Information Systems Studies, University of Sheffield School of information, Regents Court, Sheffield, UK

ABSTRACT: Knowledge and understanding of e-government infrastructure of government to business information systems remains limited. Nevertheless, identifying and addressing the barriers which hinder infrastructure in the government to business domain is still required. This paper aims to investigate barriers to infrastructure in e-government to business G2B domain to recognise the obstacles which have impact on infrastructure therefore influencing the relationship between government and business implementation. The e-Umrah government information system in Saudi Arabia was selected as the context for the investigation. A qualitative case study and an interpretive approach was used to understand the infrastructure of the e-Umrah system. Semi-structured interviews, informed by Star and Ruhleder's (1996) eight dimensions for investigating infrastructure were conducted with 43 Umrah companies. The findings indicate that the e-Umrah information system's infrastructure is affected by barriers within each of the eight dimensions of Star and Ruhleder (1996). These include lack of service integration, lack of electronic collaboration and interconnection of some entities, rigidity of standards in some operations and procedures, technical and informational breakdowns, insufficient transparency, and inefficient collaborative learning network. It is suggested when implementing G2B e-government information systems that attention be given to the technical, organizational, and social aspects of the information infrastructure.

Keywords: E-government, Implementation, Electronic services, G2B, Large scale systems

1 INTRODUCTION

The rapid evolution in the Information and Communication Technology (ICT) has greatly helped government, public and private sectors to move to an advanced way of practice. With the rise towards a digital transformation during the last decade, the government plans had set out to transform the activities from conventional way to full electronic usage (Fang, 2002). In relation to Government to Business (G2B) connection, the government has an important role in offering activities that are related to business (Gbde, 2001). The private sector is motivated to deal with digitised government and its improvement depends on the government administrative functions in utilising the information technology to implement e-government services. Certainly, the government is aware of the importance of technology development through implementing strong IT infrastructure that improves effectiveness of G2B and strengthens the relationship between the government and private sector (Gbde, 2001).

However, there is considerable emphasis on elements of obstacles and improvements. The e-government literature has emphasised generally

on some issues which should be considered in order to have successful e-government projects. These issues fall under strategy, technological, organizational, and policy (Lam, 2005; Zhang et al. 2005; Weerarakkody et al., 2008; Weerarakkody et al., 2011; GBDe, 2002; Aichholzer and Sperlich 2001; Onojaefe and Leaning 2007; Jayaradha and Shanthakumar 2003).

Strategy challenges refer to a lack of e-government goals and objectives, over milestones, ownership and governance, absence of implementation guidance, and funding issues (Lam, 2005; Zhang et al. 2005; Weerarakkody 2008; GBDe, 2002). Technology challenges include deficiencies in architecture interoperability, incompatible data standards, different security models, rigid legacy systems, and incompatible data standards (Lam, 2005; Aichholzer and Sperlich 2001; Onojaefe and Leaning 2007, Jayaradha and Shanthakumar 2003). As for the policy challenges, the major issues appear to be concerns over privacy, data ownership, and e-government policy evolution were the challenges that affecting e-government development from policy perspective (Lam, 2005; Aichholzer and Sperlich 2001; Onojaefe and Leaning 2007, Jayaradha and Shanthakumar 2003).

Challenges generally related to organisational readiness include slow pace of government reform, absence of an e-government champion, legacy government processes, lack of relevant in-house management and technical expertise. (Lam, 2005; Aichholzer and Sperlich 2001; Onojafe and Leaning 2007, Jayaradha and Shanthakumar 2003). The above section has generally outlined the major barriers surrounding e-government system. However, limited knowledge was found concerning the G2B challenges in the e-government literature which hinder our understanding of the way in which G2B implementations improves and develops. Hence, one should wonder whether the challenges previously mentioned in e-government are the same in the G2B context or additional challenges might surround G2B implementation. Therefore, this paper aims to report and discuss the barriers surrounding e-government implementation in the Government to Business (G2B) area specifically the e-Umrah system in Saudi Arabia.

2 METHODOLOGY

A qualitative case study and an interpretive approach were used to enable this research to highlight the barriers surrounding G2B in one of the e-government's large scale system. The e-Umrah government information system in Saudi Arabia was selected as the context for the investigation and involved identifying the barriers which impede the development of government to business. The idea of developing such system was to connect all government offices along with business sector which involve the Umrah activities in Saudi Arabia electronically. This is for the purpose of facilitating Umrah services for the international visitors who come from all over the globe anytime during the year to the holy cities of Makkah and Madinah to perform Umrah rituals. The system is considered a large scale system. It involved numerous ministries, government agencies, and 47 companies from the private sector (45 Umrah companies; 4 technology companies/service providers which connect Umrah companies through their interfaces to e-Umrah system). The Ministry of Hajj is the primary player and the system developer which organises, arranges, and set the rules, rights, duties, responsibilities and regulations of Umrah activities to the Saudi private service providers and Umrah companies.

Also, the case study on the e-Umrah system targeted Umrah companies' owners, managers, supervisors, IT specialists because of their broad understanding and experience in dealing with technologies, management, government's standards, regulations and other issues related to e-Umrah

system. The primary source of data for this research came from semi-structured interviews. A framework was used by Star and Ruhleder (1996) was selected. The interview questions were developed to investigate e-Umrah system in eight dimensions including embeddedness, transparency, reach or scope, learned as part of membership, links with conventions of practice, embodiment of standards, built on an installed base, and become visible upon breakdown. The collection of data took approximately eight weeks. In addition, a secondary data was another source of evidence. The information consisted of documents such as reports, diagrams, systems' manuals and guidelines from government offices and private sector. This resource helped the researcher to acquire a deep understanding of the e-Umrah system and background of the e-Umrah services which allowed the research study to achieve a wide understanding of the issues affecting government and business relations in e-government. The method of data analysis Consisted of Miles and Huberman (1994) three steps: data reduction, data display, and data reporting.

3 RESULTS AND DISCUSSION

The findings from the empirical research which was conducted to investigate the e-Umrah system revealed some important issues found in each dimension. Also, it is important to note that numerous concerns were found to affect government to business relationships.

One of the concerns included *lack of service integration and system development* which was found to have an impact on e-Umrah system and influences the use of electronic services (Installed base dimension). This is because the government consists of many departments and agencies which have different systems and databases which may not be integrated or connected with each other. These differences may be due to for example the use of different technology platforms or having inflexible legacy systems which prevent services from being fully integrated and delays the conduction of these services electronically (Lam, 2005; Ebrahim and Irani, 2006; Nurdin et al., 2011).

Another concern is *lack of electronic collaboration and interconnection of some entities among each other*. Some entities were found to be fully interconnected electronically; some were partially interconnected, while the rest were unconnected but are dealt with manually outside the electronic system (embeddedness dimension). The importance of collaboration between entities was emphasised by Chourabi and Mellouli (2011). They proposed an integrated e-government framework for service integration. The framework

considered all government stakeholders which are involved in delivering services from their sides. In addition, the framework was built to strengthen the relationship of different agencies to bridge the gap which was found to occur among agencies in the e-government project. Collaboration among different entities was one significant issue which has a substantial influence on service delivery.

Furthermore, *rigidity of standards in some services operations and procedures* was found to have rigid standards and required change in order to deliver efficient and transparent standards. This was found to have a substantial influence in the electronic system (embodiment of standards dimension). This may be due to technological and organisational challenges which influence transformation to reach full electronic practice and the negative effect of the existing legacy system over data and preventing standards to promote modern practices. (Weerakkody et al., 2006). This also means that the standards from different entities are not combined or merged to provide a set of new and modified standards that contain solutions to common problems in the e-Umrah system.

Technical and informational breakdowns were found to occur sometimes in the systems. The technical breakdown was perceived to affect the technological tools of the system (such as computers, network, website, server, etc.) where the informational breakdown was found to affect the issues related to data such as delay, failure or insufficient data. Also, absence of alternative solutions when a breakdown occurs was one of the issues found to suspend practices and hence affect business implementing tasks (become visible upon breakdown dimension). Alanezi et al. (2011) emphasised on the importance of ensuring technical functionality. This was found to provide proper service delivery to any e-government website and avoids breakdown and system failure (Alanezi et al., 2011). Furthermore service quality and technical support are considered other issues if found lacking in the e-government services, will hinder the development of e-government services and infrastructure (Rotchanakitumnuai, 2008).

Lastly, all these previous concerns were found to be the cause of an *inefficient collaborative learning network* and information exchange in the e-government system's community of practice, which affected the learning process among businesses, taking an instructive leaning process and permitting Umrah companies to shift to a situated learning process. Situated learning is self learning that is self directed and independent (Derrick, 2003). Instructive learning is a learning process which takes place by being instructed or taught by another member of the community. In the e-Umrah system, if the information is not available

in the system and was not provided by the service providers, Umrah companies gain the information through situated learning to deal with the situation and progress (Learned as part of membership dimension). Therefore, effective collaborative network is of utmost importance as the lack of communication will hinder the share of knowledge and information in addition to hindering the learning process and delaying progress and development (Leary and Fontainha, 2007). This indicates the importance of interconnectivity and communication within a community of practice.

4 CONCLUSIONS

This study revealed numerous issues which were found to influence the e-government implementation specifically in the government to business domain. The e-Umrah system is considered one of the Saudi e-government projects. The significance of this study assisted in unveiling on technical, policy, and organisational barriers which may have a considerable impact on e-government system and should be considered in G2B area.

Additionally, it is interesting to note that the research findings were validated against the existing e-government literature in tackling e-government barriers. However, this study added a head start and a new insight of knowledge to the issues surrounding G2B domain by understanding in particular the issues which should be considered in the design and implementation of G2B systems which might arise from other similar G2B system that interlink with each other and deliver interoperable services.

Thus, future research could be done investigating other G2B systems to enable researchers to compare and contrast the significance of commonality and variations among G2B systems and might lead to improving government to business relations.

REFERENCES

- Aichholzer G. and Sperlich R. (2001). Electronic Government Services for the Business Sector in Austria [Online]. Proceedings of the 12th International Workshop on Database and Expert Systems Applications, pp. 412–416. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=953096> [Accessed 10 October 2010]
- Alanezi, M. A., Mahmood, A. K., and Basri, S. (2011, September). Conceptual model for measuring e-government service quality. In Open Systems (ICOS), 2011 IEEE Conference on, pp. 411–416. IEEE.
- Chourabi, H. and Mellouli, S. (2011, June). e-government: integrated services framework. In Proceedings of

- the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times, pp. 36–44. ACM.
- Derrick, M. G. (2003). Creating environments conducive for lifelong learning. New directions for adult and continuing education, 2003(100), pp. 5–18.
- Ebrahim, Z. and Irani, Z. (2005). E-government adoption: architecture and barriers. Business Process Management Journal, Vol. 11 (5), pp. 589–611.
- Fang, Z. (2002). E-Government in Digital Era: Concept, Practice and Development. International Journal of the Computer, The Internet and Management, 10 (2), pp. 1–22.
- Gannon-Leary, P. and Fontainha, E. (2007). Communities of Practice and virtual learning communities: benefits, barriers and success factors. Elearning Papers, 5, pp. 20–29. ISSN 1887–1542.
- GBDE.org (2001). e-Government [Online]. Global Business Dialogue on Electronic Commerce. http://www.gbd-e.org/ig/egov/eGov_Recommendation_Sep01.pdf [Accessed 19 January 2009]
- GBDE.org (2002). e-Government Recommendations [Online]. Global Business Dialogue on Electronic Commerce. http://www.gbd-e.org/ig/egov/eGov_Recommendation_Oct02.pdf [Accessed 19 January 2009].
- Jayaradha, N. and Shanthakumar, C. (2003). E-Governance: Tackling the Hurdles [Online]. Tamil Internet, pp. 362–366. http://www.infitt.org/ti2003/papers/58_jayara.pdf [Accessed 15 June 2009].
- Klein, H. K. and Myers, M. D. (1999). A set of principles for conducting and evaluating interpretive field studies in information systems. MIS quarterly, pp. 67–93.
- Lam, W. (2005). Barriers to e-government integration. Journal of Enterprise Information Management, 18, pp. 511–530.
- Miles, M. B. and Huberman, A. M. (1984). Qualitative data analysis: a sourcebook of new methods; Qualitative data analysis: an expanded sourcebook (2nd ed.). Thousand Oaks, CA: Sage.
- Neuman, W. L. (2004). Basics of social research: Quantitative and Qualitative Approaches. Pearson.
- Nuridin, N., Stockdale, R. and Scheepers, H. (2011). Understanding organizational barriers influencing local electronic government adoption and implementation: the electronic government implementation framework. Journal of theoretical and applied electronic commerce research, 6(3), pp. 13–27.
- Onojaef, D. and Leaning, M. (2007). The Importance of Partnerships: The Relationship between Small Businesses, ICT and Local Communities. Issues in informing science and information technology, 4, pp. 725–736.
- Rotchanakitumnuai, S. (2008). Measuring e-government service value with the E-GOVQUAL-RISK model. Business Process Management Journal, 14(5), pp. 724–737.
- Star, S. L. and Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. Information systems research 7.1, pp. 111–134.
- Weerakkody, V., Dhillon, G., Dwivedi, Y. and Currie, W. (2008). Realising Transformational Stage E-Government: Challenges, Issues and Complexities [Online]. AMCIS Proceedings. Paper 181. <http://aisel.aisnet.org/amcis2008/181/> [Accessed 06 November 2010].
- Weerakkody, V., El-Haddadeh, R. and Al-Shafi, S. (2011). Exploring the complexities of e-government implementation and diffusion in a developing country Some lessons from the State of Qatar. Journal of Enterprise Information Management. 24(2), pp. 172–196.
- Weerakkody, V., Baire, S. and Choudrie, J. (2006). E-government: the need for effective process management in the public sector. In System Sciences, 2006. HICSS '06. Proceedings of the 39th Annual Hawaii International Conference. IEEE. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1579436&tag=1.
- Zhang, J., Dawes, S. and Sarkis, J. (2005). Exploring stakeholders' expectations of the benefits and barriers of e-government knowledge sharing. Journal of Enterprise Information Management, 18(5), pp. 548–567.

The effectiveness of an educational website for promoting design skills and use of educational blogs by teachers of secondary education in Saudi Arabia

Mohammad Ahmad Saeed Al Mozaini

Department of Education, King Abdul-Aziz University, Jeddah, Saudi Arabia

ABSTRACT: The main purpose of this article is to investigate the effectiveness of an educational website for promotion of design skills and use of educational blogs by teachers of secondary education as the educational blogs are considered one of modern technology tools that facilitate the teacher to present his/her educational material in an interesting and effective manner, and provide the teacher to communicate with his/her students from distance and reveal their skills of writing, editing, and compiling the information. So, this article mainly has relied on the use of two types of tools to measure the sample skill, aimed to design and use of educational blogs; which are note card and pre training test. The two tools were employed in a sample of tribal test, then the training of the sample was held through website designed especially for training of the design skills and use of educational blogs, after that the research has applied the note card and pre-training test to measure the effectiveness of website in promoting of sample's skill in design and use of educational blogs. The main finding of this work prove that there are some significant statistical differences at the level of 05.0 among average class of teachers revealed in note card, for pre and post training that is in favor of post training interest. It reveals the effectiveness of the educational website for promotion of design skills and the use of educational blogs by teachers of secondary education.

Keywords: Electronic Education, Learning Management System, Educational Blogs, Computer Based Education, Distance Education, Second Web Generation, Wiki and Social Favorite

1 INTRODUCTION

The contemporary world has been witnessing radical change in the fields of communication and information, as well as in the field of education; in terms of teaching methods, and variety of information sources and means of acquiring knowledge. Perhaps the most prominent of these ways and means was discovered, for us, after the entry of computer and internet, with all its possibilities, capabilities and with full force in the field of education, till it has become a key partner today in educational process. It is very difficult for the contemporary education system to remain isolated from this technology, because it has become a very essential means for preparing the lessons, and presentation, even for entire educational segment, that is called "Electronic Education".

With the rapid technological development in the field of computer and internet and with the emergence of new technologies for internet services, a number of modern concepts have been crystallized in computer-based education through the development of software and systems, that are called "Second generation of web", this term was first used in

a conference of same name (Web 2.0 conference) held by O' Reilly Company. The term is an outcome of brainstorming in a meeting which was held on the sideline of the conference organized by O'Reilly Company and Media Live International, where the Web 2.0 was defined as "a group of websites, services and applications that bears certain features of (O' Reilly, 2005)"; some are given below:-

- To provide high level of interaction with the user: It means, while using the web applications, a user should realize full interaction with the content, and find himself/herself capable to add and modify in similar way to the work, on his PC (personal computer).
- Participation of the user in the content: In the past, the web was a platform for reading only, so the content available on internet was input by particular people working under companies, or organizations or government, and the common user of internet had no chance to add any content. But nowadays it has become possible that a user can add and modify the web contents—that could be done easily—and the user has become the key focus in the process of enriching

the content, thus the web applications, such as blogs and Wiki, have contributed at large scale to form a Read/Write Web, after changing from a platform for reading only.

During working in education department of Jeddah province as a supervisor of electronic education, and participating in various meetings and seminars, and following up international conferences pertaining to electronic and distance education, as well as by soliciting the views of teachers during supervisory visits in the schools, the researcher came across to know that the majority of the teachers are not well familiar with the design and usage standards of educational blogs. And in the lights of above mentioned facts in the forward, the research problems could be summarized in order to identify the effectiveness of educational website for promotion of design skills and use of educational blogs by teachers of secondary education.

2 RELATED WORKS

In the following paragraphs, we shall present some important works that are done in the field of Web 2.0 applications especially in educational process.

- Study of Abu Anaqa, 2014 entitled “Electronic Education in the Age of Web 2.0”; this is a field study regarding use of educational blogs in the field of Teaching Library Science in Mentouri University—Constantine). This study is aimed to find out the reality of the use of educational blogs in teaching library science, through knowing the views of teaching faculty members of the Department of Library Science, University of Mentouri—Constantine, regarding the use of this technology in teaching field.
- Study of Abdul Basit, 2013 entitled “The reality of essential opportunities of the use of electronic blogs in teaching field among human sciences’ teachers of KSA”. This study aimed to specify the personal and administrative reality as well as basic educational opportunities regarding the use of electronic blogs, in teaching field among human sciences’ teachers of Kingdom of Saudi Arabia. The study concluded that the extent of personal actual ability of using e-blogs, among social science teachers is very low, although the result of the study declared that the plurality and diversity of essential educational opportunities are significantly available for the social science teachers, as well as the results of the study also declare there are no statistically differences found between male and female teachers regarding administrative and basic educational opportunities assessment, for using e-blogs in teaching of social sciences.

- Study of Imran, 2012, entitled “Effectiveness of the use of educational blogs in teaching of Geography, for acquiring knowledge and developing geographical research skills and motivation among students of first year students of secondary school”. In this work, the researcher has tried his best to measure the effectiveness of the use of educational blogs in Geography teaching on the scale of acquiring knowledge and boosting geographical research skills. The results of the study further indicate that there are statistically significant differences found between those who use educational blogs and those who do not use them, and that is in the favor of the first group.
- Study of Muslim, 2011 entitled “Impact of Biology Science teaching through educational blogs, on developing learning motivation and boosting educational communication skills among first year students of secondary school”. This study aimed on measuring the impact of the use of educational blogs on developing the learning motivation and boosting educational communication skills among first year students of secondary school. The study concluded that the use of educational blogs provides the students high level of motivation for participation, particularly among the students who feel shy in class room. Also that educational blogs have a number of graphics, diagrams, video clips which increase the rate of interest among the students. As well the educational blogs also could be used for the subject exercise (e.g. Bio subject), hence the educational blogs have turned into a comprehensive source of subjects exercises, students may consult them later on, and it also may enhance their interests.
- Study of Al Misri, Salwa, 2011 entitled “Effectiveness of use of the Educational Blog in increasing the knowledge of computer subjects among primary students and their orientation toward education science”. This study aimed to measure the effectiveness of use of educational blogs to increase the extent of understanding the computer subject. It was found in the conclusion of the study that there is a great and interesting impact of the use of educational blogs in boosting knowledge among primary students, and it has played a vital role through presenting important meanings. At the same time, significant improvements were also found in the students’ attitudes toward computer subject.

3 PROBLEM STATEMENT

Working in Education Department of Jeddah province as a supervisor of electronic education,

and participating in various meetings and seminars, and following up international conferences pertaining to electronic and distance education, as well as by soliciting the views of teachers during supervisory visits in the schools, the researcher came across to know that the majority of the teachers are not well familiar with the design and usage standards of educational blogs. And in the lights of above mentioned facts in the forward, the research problems could be summarized in order to identify the effectiveness of educational websites for promotion of design skills and use of educational blogs by teachers of secondary education, where the study attempts to answer the following questions:-

- What are the standards and parameters of educational blogs' designing?
- What is the fundamental skills required for educational blogs' designing?
- What is the effectiveness of (internet based) educational website for the developments of educational blogs among secondary school teachers?

4 METHODOLOGY OF SOLUTION

This study uses an experimental method, in terms of relying on the first design of experimental method e.g. pre and post test approach, through using a single experimental group. This study contains two types of basic variables; one is independent variable that is 'internet based educational website', and another one is dependent variable that is 'skills of educational blogs' designing'. The researcher assesses the value of the phenomenon before holding the experiment, and later on uses to assess the value again after the analyzing (of the sample) of the experimental variable. According to the researcher, the thing which differentiates between the two assessments is the influence of variable experiment on the phenomenon during the course of discussion, which is shown in the following table1:-

Table 1. Variables experiments.

Experiment of the sample	Independent Variable (internet based educational web site) Dependent Variable (Skill of Educational Blogs' designing)
Q.1	Assessment before the experiment (Pre test)
Q.2	Assessment before the experiment (Post test)

5 THE MAIN FINDINGS AND NUMERICAL RESULTS

In order to examine the hypothesis which states that: there are some significant statistical differences at the level of 05.0 among average class of teachers revealed in note card, for pre and post training that are in favor of post training interest, it should answer the following research questions?

Answer of Q.1: What are the standards and parameters of educational blogs' designing?

An achievement test, containing of standards and parameters of educational blogs' designing, was held. Table 2 shows the descriptive statistics of the both pre and post exams of the teachers in the achievement test.

Table.2 reveals the number of sample is 20, and the average number of teachers attended the pre test event is 17.55%, with a deviation rate of 3.4%, while in the post test the average of their presence was found about 23.6%, with 3.6% deviation rate. On other hand, Table.3 demonstrates the significant differences between averages level of teachers appeared in pre and post exams for the achievement test.

By extrapolating the results of the above Table.2, it becomes clear that the independency level is 19%, while relevant sample test value tolls 12.5-%, and as for the SIG test is concerned, it is 0.00, which is less than the level of 0.05, thus it would be statistically significant, and indicates that there are some significant statistical differences at the level of 05.0 among average class of teachers revealed in note card, for pre and post training that is in favor of post training interest.

Answer of Q.2: What are the fundamental skills required for educational blogs' designing?

The note card for performance skills was prepared, which consists of basic skills for design and assessment of educational blogs. The Table.4 shows the descriptive statistics of the both pre and post exams of the teachers in the note card.

From Table.4, it is shown that the number of sample is 20, and the average number of teachers

Table 2. Descriptive statistics of the two pre and post exams of teachers in achievement test.

Exam/Test	Number of examinees	Average %	Deviation %
Pre Test	20	17.55	3.4
Post test	20	23.6	3.6

Table 3. Significant differences between averages level of teachers appeared in pre and post exams for the achievement test.

Exam/Test	Number of examinees	Average %	Deviation %	Degree of independency	Test-T	Indication
Pre Test	20	17.55	3.4	19	12.5-	There are statistically differences between the two tests, at a level of 0.05%
Post Test	20	23.6	3.6			

Table 4. Descriptive statistics of the both pre and post exams of the teachers as reflected in note card.

Exam/Test	Number of examinees	Average %	Deviation %
Pre Test	20	4.85	3.36
Post test	20	55.4	8.69

Table 5. Significant differences between averages level of teachers appeared in pre and post exams for the not card.

Exam/Test	Number of examinees	Average %	Deviation %	Degree of independency	Test-T	Indication/SIG
Pre Test	20	4.85	3.36	19	29.122-	There are statistically differences between the two tests, at a level of 0.05%
Post Test	20	55.4	8.69			

attended the pre test event is 4.85%, with a deviation rate of 3.36%, while in the post test the average of their presence was found about 55.4%, with 8.69% deviation rate. While the following Table.5 demonstrates the significant differences between averages level of teachers appeared in pre and post exams for the note card.

By extrapolating the results of the above Table.5, it is clear that the independency level is 19%, while relevant sample test value tolls 29.122-%, and as for the SIG test is concerned, it is 0.00, which is less than the level of 0.05, thus it would be statistically significant, and indicates that there are some significant statically differences at the level of (05.0) among average class of teachers revealed in note card, for pre and post training that is in favor of post training interest.

Answer of Q.3: What is the effectiveness of internet based educational website for the developments of educational blogs?

While answering of the two previous questions and the evaluation of table no.3 and 4, the findings of which support the post training test, the answer of the above third question is found, as well as the hypothesis also could be accepted.

6 CONCLUDED REMARKS

This study was aimed on the assessment of effectiveness of internet website to develop skills of

design and use of educational blogs by teachers of secondary education, and the research reached at a conclusion that an appropriate training of the teachers for the use and designing of educational blogs through a web site, will be resulted in a number of benefits and academic advantages, which may boost the skills of teachers to use and design educational blogs. As a matter of fact, websites have reached their ultimate goal of designing, and through this a number of teachers were able to boost their skills for using and designing the educational blogs, as well as, by this they came across to know the standards of blogs designing, in a scientific manner. Thus, the researcher recommends to take advantages of modern technology, especially the advantages of websites, which can be used to train the teachers and develop their skills in several spheres pertaining to their educational mission where these websites can help to achieve a number of goals; most importantly saving time, and efforts, and training a number of group without any limitation of number, or time or place. Moreover, these sites can also help to reduce the cost of training, and can contribute actively to the professional growth of the teachers. The researcher also recommend to take advantages of technical tools offered by internet technology, especially which has been facilitated by social sites, e.g. blogs and wikis etc., which provides very attractive services and make easier to communications between teachers and students and allow them to teach and learn at any time and any place.

REFERENCES

- Abdul Basit, H.M.A. (2013), Al-Waqe w al-Furas al-Lazimah Li- Istikhdami al-Modawwanati al-Electroniyyah Fi al-Tadrees Lada Mo'allimi w Mo'allimati al-Uloom al-Ijtimaiyyah, KSA, Journal of Education and Psychological Sciences, Bahrain, Vol.14, Issue.2, pp. 369–394.
- Abu Anaqah, S. (2014), Al- Taleem al-Electroni Fi Asr al-Web 0.2: Dirasah Maidaniyyah Ala Istikhdami al-Modawwanati al-Talimiyyah Fi Tadreesi Ilmi al-Maktabati, Mentour University, Constantine, Jordanian Magazine of Libraries and information, Jordanian Association for Libraries and Information, Jordan, Vol.39, Issue.3, pp. 11–35.
- Al-Misri, C.F.M. (2011), Failyatu Istikhdami Modawwanatin Talimiyyatin Fi Ziyadati Tahsili Tullabi al-Marhalati al-I'dadiyyati li-al Mafaheemi al-Mojaradati bi-Maddati al-Combuter w al-Ittijahi Nahw al-Maddati, Education Journal, Egypt, Vol.19, Issue.4, pp. 171–228.
- Ayad, F.E. (2015), Failyatu Modawwanatin Talimiyyatin Li-Masaqi Taqniyyati al-Tadreesi Fi Tanmiyyati al-Tahseel- Al-Marfee w Usloobu Al-Ta'allumi al-Ameeqi w Darajati Qubooli Lada Talibati Jamiati Al-Aqsa, Journal of Education and Psychological Sciences, Bahrain, Vol.16, Issue.3, pp. 517–563.
- Imran, Kh.A'.A.M. (2012), Failyatu Istikhdami al-Modawwanati al-Talimiyyah Fi Tadreesi al-Joghrifiyyah Ala al-Tahseeli al-Ma'rifi w Tanmiyyati Maharati Al-Bahs al-Joghrifi w al-Dafiyyati li-at-Ta-Allumi Lada Tul-Labi al-Saffi al-Awwali al-Sanaviy, Education Journal, Egypt, Vol.31, pp. 353–425.
- O'Reilly, Time (2005), What Is Web 2.0, Design Patterns and Business Models for the Next Generation of Software. Retrieved, Dec 25, 2015, from: <http://oreilly.com/web2/archive/what-is-web-20.html>.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Using machine learning technique towards personalized mobile flyer

M.A. Razek & H.G. Bardessi

Research and Development Department, Deanship of Distance Learning, King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT: The exponential growth of mobile communication technologies provides marketing with opportunity to connect with consumers directly on their mobile phones beyond traditional and digital media. Nowadays, advertising via mobile devices is very interested, however, it still demand additional investigation. The underlying paper discusses how to determine the nearest consumers around the target user to personalize a suitable flyer. The interested items based on consumers' similarities are collected. Then the collection is used to personalize a suitable flyer to the target customer. The paper builds an algorithm called Consumers Similarity Algorithm (CSA) that measures similarities among consumers' profiles and return a list of similar users. An investigation was done to detect the impact of the CSA on the classification used by K-Nearest Neighbor method. An experimented was conducted and the results were discussed using kNN-CSA and without it. Our experiments show that kNN gives a high considerable performance when merged with CSA than its performance alone.

1 INTRODUCTION

The fast developing in using multimedia on Mobile devices assist on moving marketing from e-marketing to m-marketing. Nowadays, most of the vendors straight connect with their clients anytime and anywhere. Geumhwan et al. [1] mentions that mobile advertisement is progressively used among many applications. With the growing mobile app markets, mobile advertising has become an offensive key factor to force economic in recent years [2].

This paper displays how to customize and adapt a mobile flyer depended on the user's needs. It uses two ways to improve the personalization of the flyer 1) customizing the flyer based on consumers' needs, and similarities among user profiles, 2) refining the personalization by improving flyer categories using k-nearest neighbor classification. The paper sees the flyer as a set of visual information about items that best fit an intended information that the user needs. Following [3] in representing the classes, the flyer contains a set of categories, each category holds visual information about some items that fall within the interested of the consumers' needs.

To overcome the above challenges, the paper presents solutions for, what is the methodology to adapt a collection of categories based on buyers' needs? What is the methodology to be used for selecting similarity buyers related to the target buyer? How to select a right item to create a personalized flyer?

The reset of the paper is organized as follows. In the next section, we elaborate related work and other similar research results while section 3

describes methodologies to compute user preferences. Section 4 provides more details about using K-Nearest Neighbor (KNN) technique for classifying items classes. Section 5 shows dataset, results, and analysis of our experimentations. Finally, section 6 concludes the research results.

2 RELATED WORK

The mobile market is growing rapidly, because the number of mobile terminals is continuously increasing in the world [4].

Advertising is an essential component of marketing communications and its industry has rapidly adapted from bill board to internet [5]. Leppäniemi et al. [6], considers mobile marketing as "the distribution of any type of message or promotion that brings value to the consumer while improving benefits for the company". Bauer [7] considers that mobile advertising represents a particular case of the application of the TAM, that is, an innovation of use: "the communication of a content via a mobile media can only be efficient if the consumers allow the regular reception of advertising messages on their mobile phones".

Pihlström and Brush [8] examined how the information and entertainment mobile content service users influences repurchasing and paying a price premium. They have used a sample of 579 mobile service users. The findings recognized differences between service user groups, and discussed for the use of diverse marketing strategies for entertainment and information mobile services.

This approach helped authors to understand the behaviors of individuals for post-purchase.

On the other hand, recommended systems are almost using collaborative filtering to predict a user's preference on users' rating items. In the recommended system, there is differentiated between user preferences and item preferences, while user preferences specify the perceived distinct change among users, item preference shows the observed preference for each item compared to the overall average [10].

There are many advantages for using K-Nearest Neighbor Technique (kNN) in finding the nearest similar users to the current one. However, KNN method is very responsive to the select of the similarity function that is used to compare between users preferences; it requires a big storage to store the instances [11].

Our technique for finding similarity between users by using Pyramid Collaborative Filtering Model (PCFM) [12] along with K-Nearest Neighbor Technique [13], rather than using the heuristic similarity measure that focuses on improving recommendation performance under cold-start conditions where only a small number of ratings are available for similarity calculation for each user [11]. The key drivers for advertisers to include Mobile Personalized Marketing [14] in their media plans are as follows, 1) Personalization can be achieved, 2) Differentiation in message delivery, 3) Mass market can be widely addressed, 4) Resources can be utilized better.

On the other hand, Drossos et al. [15] conducted another study included some factors such as mobility, interactivity or personalization and found that customer connection, motivation level, and manner to SMS flyer affected on buying intents.

Obviously, what needed is a conceptual framework from which to interpret these diverse findings. Leppäniemi and Karjaluoto [16] investigated some issues effecting customers' readiness to admit mobile flyer. They suggest that customer readiness to admit mobile flyer will be subject on a user-friendly technology, a good mobile marketing, a successful personalization of messages, monitoring guarantees of privacy, and relevant information.

The next section sheds light on the methodology of calculating the user preferences.

3 USER PREFERENCES METHODOLOGY

In 2010, a study conducted by ABI Research [17] reported that over one billion new mobile handsets were shipped in 2009, with another 1.2 billion projected to ship in 2010. The communication method used in Mobile is one of the best key issues that would lead a media planner to select mobile media for the communications strategy between vendors

and their consumers [18], [19]. However, Yunos et al. [20] mentioned that flyers would stand up to some technical provocations such as varying configuring in the flyers, and mobile screen. These provocations have been overcome with the new version of iPad and new Smartphone such as Galaxy S released in June 2010, which adopt HTML5 formatting and screen with high resolutions.

3.1 User preferences representation

The flyer consists of m categories; each category belongs to a specific domain such as electricity, foods, computer, clothes, etc. the flyer formatting used multimedia such as picture, images, HTML5, or pdf. In general, suppose that C is a finite set of m categories of items, and each category has n items: $C = \{C_k\}_{k=1}^m$,

$$C_k = \{d_{ki} \mid i = 1, \dots, n\} \quad (1)$$

Therefore, the categories is a $m \times n$ -tuple,

$$\begin{bmatrix} d_{11} & \dots & d_{1n} \\ \vdots & d_{ij} & \vdots \\ d_{m1} & \dots & d_{mn} \end{bmatrix} \quad (2)$$

where d_{ij} represents the j th chunk (item) in the i th category. Consequently, the following Φ_k ($m \times n$ -tuple) represents the user's degree of interest for all categories for a user u_k .

$$\Phi_k = \begin{bmatrix} \phi_{11}^k & \dots & \phi_{1n}^k \\ \vdots & \phi_{ij}^k & \vdots \\ \phi_{m1}^k & \dots & \phi_{mn}^k \end{bmatrix} \quad (3)$$

Where

$$\phi_{ij}^k = \frac{f(u_k, c_{ij})}{\sum_{i=1}^n \sum_{j=1}^m f(u_k, c_{ij})} \quad (4)$$

$$\sum_{i=1}^n \sum_{j=1}^m \phi_{ij}^k = 1$$

where $f(u_k, c_{ij})$ represents the frequency number, when a user u_k selects a chunk c_{ij} .

4 ITEM CREDIBILITY

In case, a user just selected items and he/she did not buy any of them, a credibility matrix overcome this problem. The matrix represents the interested items based on the purchases. Accordingly, we can define the credibility matrix of the user as follows:

$$\Delta_k = \begin{bmatrix} \lambda_{11}^k & \dots & \lambda_{1n}^k \\ \vdots & \lambda_{ij}^k & \vdots \\ \lambda_{m1}^k & \dots & \lambda_{mn}^k \end{bmatrix} \quad (5)$$

$$D(u_k, C_j) = \sum_{l=1}^n \lambda_{jl}^k \quad (6)$$

$$\text{where } \lambda_{ij}^k = \frac{S(u_k, c_{ij})}{\sum_{i=1}^n \sum_{j=1}^m S(u_k, c_{ij})}$$

$$\text{where } \sum_{i=1}^n \sum_{j=1}^m \lambda_{ij}^k = 1$$

where $S(u_k, c_{ij})$ represents the frequency number when user u_k buys the chunk c_{ij} . Accordingly, we can compute the credibility for user u_k for each category C_i as follows:

$$\Omega_k = \Theta_k + \Delta_k = \begin{bmatrix} \omega_{11}^k & \dots & \omega_{1n}^k \\ \vdots & \omega_{ij}^k & \vdots \\ \omega_{m1}^k & \dots & \omega_{mn}^k \end{bmatrix}, \quad (7)$$

$$\text{where, } \omega_{ij}^k = \phi_{ij}^k + \lambda_{ij}^k, i=1, \dots, n \& j=1, \dots, m$$

$$\text{where, } \omega_{ij}^k = \phi_{ij}^k + \lambda_{ij}^k, i=1, \dots, n \& j=1, \dots, m$$

Based on formula (7), we can define an interested set Ψ_i^k for each category C_i for each user u_k :

$$\Psi_i^k = \{\omega_{ij}^k \mid i=1, \dots, n\} \quad (8)$$

Accordingly, we can compute the interested of the user u_k for each category C_j :

$$I(u_k, C_j) = \sum_{i=1}^n \Psi_{ij}^k \quad (9)$$

5 FLYER PERSONALIZATION

Our approach is to looking for the users who are similar to the intended user in order to personalize his flyer. To overcome that this section describes two approaches to predicting those users: Consumers similarity algorithm and k-nearest neighbor method.

5.1 Consumers similarity algorithm

In this subsection, we determine the specifications that effects on the similarity between users. We collect some characteristics, which are common between customers and the current cus-

tomers u_k , such as interesting on the same items and similar needs using Consumers Similarity Algorithm (CSA). Accordingly equation (8), we can represent the interested matrix for a user u_g and u_t :

$$\Omega_g = \begin{bmatrix} \omega_{11}^g & \dots & \omega_{1n}^g \\ \vdots & \omega_{ij}^g & \vdots \\ \omega_{m1}^g & \dots & \omega_{mn}^g \end{bmatrix}$$

$$\Omega_t = \begin{bmatrix} \omega'_{11} & \dots & \omega'_{1n} \\ \vdots & \omega'_{ij} & \vdots \\ \omega'_{m1} & \dots & \omega'_{mn} \end{bmatrix}$$

For each matrix, we generate an on-off matrix Γ . The elements of the on-off matrix is either one or zero.

$$\Gamma(\Omega_g) = \begin{bmatrix} \phi_{11}^g & \dots & \phi_{1n}^g \\ \vdots & \phi_{ij}^g & \vdots \\ \phi_{m1}^g & \dots & \phi_{mn}^g \end{bmatrix},$$

$$\Gamma^{-1}(\Omega_t) = \begin{bmatrix} \phi'_{11} & \dots & \phi'_{1n} \\ \vdots & \phi'_{ij} & \vdots \\ \phi'_{m1} & \dots & \phi'_{mn} \end{bmatrix}$$

$$\text{where, } \phi_{ij}^g = \begin{cases} 1 & \omega_{ij}^g > 0 \\ 0 & \omega_{ij}^g = 0 \end{cases}$$

$$\text{and } \phi'_{ij} = \begin{cases} 1 & \omega'_{ij} > 0 \\ 0 & \omega'_{ij} = 0 \end{cases}$$

The similarity between two people u_g and u_t is given by:

$$S(u_g, u_t) = \frac{1}{mn} \sum_{i=1}^n \sum_{j=1}^m \phi_{ij}^g \times \phi'_{ij} \quad (10)$$

Step 1: Consumers Similarity Algorithm (u_g, u_t)

- For each customer u_g, u_t ,
 - Compute $S(u_g, u_t)$
 - For k if $S(u_g, u_t) \geq 0.5$ return u_g, u_t are similar for a concept.
-

Based on formula (10), we can define a set of users' similarity for a user u_k :

$$U^k = \{u_i \mid S(u_k, u_i) \geq 0.5, \forall i=1, \dots, h\} \quad (11)$$

Accordingly and based on formula (9), we can compute the interested degree for all users, which are similar to the user u_k for each category C_j :

$$I(U^k, C_j) = \frac{1}{h} \sum_{i=1}^h \sum_{l=1}^n \Psi_{jl}^h \quad (12)$$

5.2 K-Nearest Neighbor classification

The K-Nearest Neighbor (KNN) technique is a simple and very intuitively appealing method to address classification problems and it is a case-based learning method, which keeps all the training data for classification [10] and [12].

We use KNN as another technique rather than in subsection 3.3 to determine the K the k-nearest users around u_k . Accordingly, KNN classifier views a flyer as a set of conditionally independent items $U = \{u_1, u_2, \dots, u_m\}$, which is represented as a collection of m users, and $u_k = \{u_{i1}, u_{i2}, \dots, u_{in}\}$ be the k nearest neighbors of u_k . The testing examples are represented by U_i with m_i points; x_o is an arbitrary testing examples fact, and $U_0 = \{u_{01}, u_{02}, \dots, u_{0k}\}$ covers its k nearest neighbors from training examples, with labels $\{l_1, l_2, \dots, l_k\}$. Supposed that there are r classes in the examples collection $\Gamma = [\gamma_1, \gamma_2, \dots, \gamma_r]$.

The class label assigned to a test example is determined by the majority vote of its k nearest neighbors.

$$Class = \arg \max_k \sum_r \delta(l_r, \gamma_k), \quad (13)$$

where δ is the Kronecker delta

$$\delta(l_r, \gamma_k) = \begin{cases} 1 & l_r = \gamma_k \\ 0 & \text{otherwise} \end{cases}$$

And we use the Euclidean distance function $d(z, w)$ between two examples z and w :

$$d(z, w) = \sqrt{\sum_{i=1}^m (z_i - w_i)^2} \quad (14)$$

Following NKK [12] and [13], the input of the NKK algorithm is requested item d_v^k , and the output would be requested concept u_k .

5.3 Personalized flyer for a user

To create a flyer for a user u_k , firstly we determine the order of the categories, and then rearrange the contents of every category. For determining the order of the categories, we sort as descending order the set $C = \{C_k\}_{k=1}^m$ based on the value $I(U^k, C_j)$ in formula (11). Each category is represented by a set $C_j = \{d_f^k \mid f = 1, \dots, n\}$.

Where d_f^k presents the frequency of item d_f , which belongs to category C_j for a user u_k :

$$d_f^k = \frac{1}{3} \left(\frac{n_f}{F_{\max} \times N} + I(U^k, C_j) + I(u_k, C_j) \right) \quad (15)$$

where n_f represents the number of items of the item d_f^k had been sold, N represents the total number of receipts, and F_{\max} represents the maximum number of items for an item had been sold, and $I(u_k, C_j)$ represents the interested of the user u_k for each category C_i represent sum of user interested on item v :

$$F_{\max} = \text{Max}_{v=1, \dots, n} \{n_f\}$$

6 EXPERIMENTAL METHODS AND RESULTS

The interested items based on consumers' similarities are collected. Then they are used to personalize a suitable flyer to the target customer. To find the similarities between the users, we try to find the similarity between the items selected by the users. To do that, we used two methods: K-Nearest Neighbor Classification alone, and K-Nearest Neighbor Classification along with Consumers Similarity Algorithm. An investigation was done to detect the impact of the CSA on the classification used by K-Nearest Neighbor method.

The data set provides information on selecting items from electronic advising for fours items categories: Laptop, PC, Smart-phone, and Tablet. Each category has four items represented 4-tuple, as follows;

- Laptop = [Dell, Hp, Apple, Toshiba],
- PC = [Dell, Hp, Apple, Toshiba],
- Smart-phone = [iPhone, Galaxy, LG, HTC],
- Tablet = [iPad, Galaxy Note, iPad mini, Galaxy mini].

The number of the item selected by the user represents each item. The data set presents 300 examples, which collected using 300 users (160 female). Each user is invited to one attempt to choose his favorite item from each category electronically. Table 1 shows the distribution of the selected categories. The effectiveness of category for the two classifiers methods is evaluated by the standard precision, recall, and F1 measure. Recall is the ratio of correct positive predications by the system divided by the total number of positive examples. Precision is defined to be the ration of correct positive predications by the system divided by the total number of positive predications. F1 measure contains two components recall and precision in the following way:

Table 1. Dataset characteristics.

Class name	Number of labeled data
Laptop	75
PC	75
Smart-phone	75
Tablet	75
Total instances	300

$$F_1 = \frac{2 \times recall \times precision}{(recall + precision)}$$

The second experiment, the dataset was converted using equations from formula (1) into formula (8) to calculate the value of each slot. Accordingly, we applied Consumers Similarity Algorithm (CSA) to find the similarity users and then get their corresponding examples and apply again K-Nearest Neighbor classification. The results and its analysis is shown in the next subsections.

6.1 Results and discussion

This section presents our results and analysis of the impact of Consumers Similarity Algorithm on the performance of K-Nearest Neighbor classification. We used two-fold cross validation by conducting the two methods (kNN only, and kNN-CSA) two times and then calculating the average of the two performances as result.

As it can be seen, in almost all experiments for the two methods, the performance of k-NN depends on the number of nearest neighbor k and the number of the features used. Figure 1 shows the precision and recall of using kNN classification only. The results of precision and recall compared against the number of the features. Four collections have been taken to conduct the experiment: 25%, 50%, 75%, and 100% of the whole dataset for each category.

Figure 2 shows the precision and recall of using K-Nearest Neighbor classification with CSA for PC Category. From the experimental results, it is clear that KNN-CSA shows a significant classification performance improvement outperforms the k-NN algorithm. The time taken to build model in KNN is around 0.02 seconds, however, the time taken by KNN-CSA is 0.004, which is better. The correctly classified instances by KNN is 66% out of 300 (the total number of instances) however, the correctly classified instances by KNN-CSA is 92.7% which explain the value of mean absolute error for KNN-CSA is 0.0532 and is equal to 0.3566 in KNN.

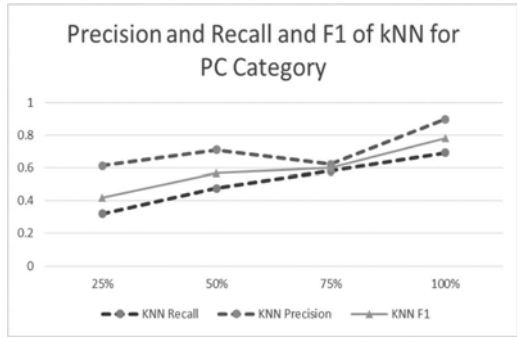


Figure 1. Precision and recall of KNN for PC category.

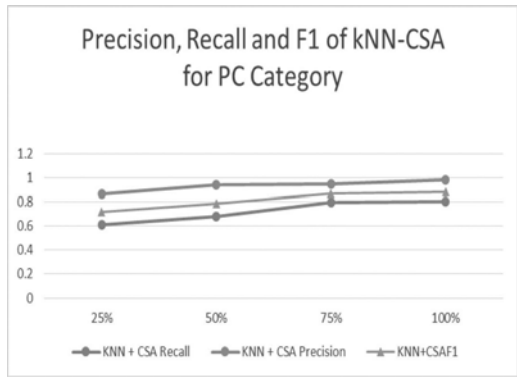


Figure 2. The impact of CSA on the recall of KNN for PC category.

For example in the instances clustering in the categories, we notes that the KNN-CSA classified 69 instances out of 75 in the Tablet-category, and 67 instances out of 75 in Laptop-Category. However, the worst case was at Smart-phone Classification where the KNN-CSA classified 50 instances out of 75. As it can be seen in Figure 2, the F1 measure is clearly growing up in almost all experiments, conversely, in Figure 1, the F1 measure in k-NN drops down at 75% point. Accordingly, the k-NN precision falls down at 75% point. We can see that both F1 measure and its precision values rise significantly for 100% point.

The overall F1 measure for KNN-CSA and kNN achieved their best performance when the collection is 100%. On average, the performance of KNN-CSA is at least 20 times faster than that of kNN algorithm.

Figure 3 shows the interchange between precision and recall, and the resulting F1 for kNN for all dataset. Note that Smart-Phone category has

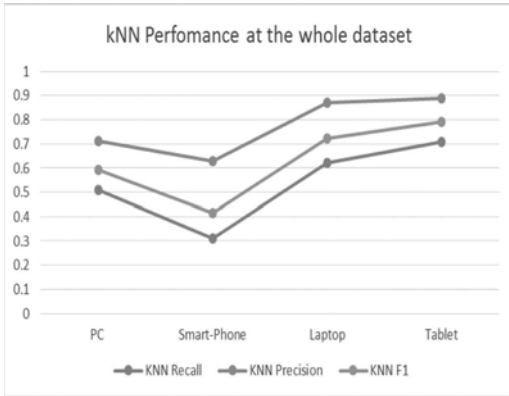


Figure 3. The kNN performance at 100% of the dataset for all categories.

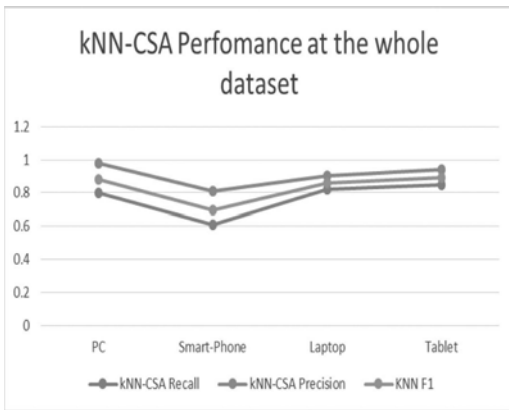


Figure 4. The kNN-CSA performance at 100% of the dataset for all categories.

lower values for three measures; however, Tablet category has greatest values for three measures.

Figure 4 shows the comparative results of the kNN-CSA performances for 75 examples for each group. Note that the results carefully simulate the same performance of kNN method at the lowest performance that records at Smart-Phone category, yet, it records the best performance at PC category.

Figure 5 presents the impact of using CSA with kNN. Notably, the F1 values of kNN-CS are better than kNN alone at all points. The largest difference between them occurs in PC category with (0.29), where the lowest difference was at Tablet category with difference (0.10). The result shows that the kNN-CS method reaches considerable F1 value if compares to the kNN method. That

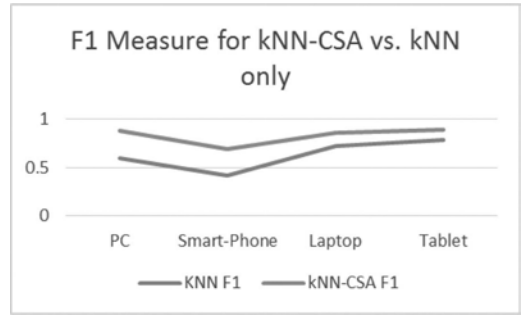


Figure 5. F1 measure for kNN-CS vs. kNN-CSA.

is because kNN-CS provides more convergent examples than kNN.

7 CONCLUSION

In this paper, we have proposed an algorithm to help k Nearest Neighbor classification to personalize a suitable flyer to consumers by choosing the most similar neighbor users. The paper showed the methodology of the algorithm and presented an investigation of two approaches: kNN with Consumers Similarity Algorithm (CSA), and kNN alone.

An experiment was conducted and its results were discussed. The analysis of the results was based on data from the experiment of 300 consumers who were asked about their interested items. The key finding of the results was proposing where the score of F-Measure for kNN with CSA indicates that there is a high performance of the construction of the classification model.

REFERENCES

- [1] Geumhwan Cho; Junsung Cho; Youngbae Song; Hyounghick Kim, An Empirical Study of Click Fraud in Mobile Advertising Networks 2015 10th International Conference on Availability, Reliability and Security (ARES), pp. 382–388 (2015).
- [2] Jinghua Jiang; Zhenkui Shi; Xingliang Yuan; Cong Wang; Xiaolin Gui, Towards secure and practical targeted mobile advertising. 2015 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), pp. 79–80 (2015).
- [3] Razeq A.M., Credible Mechanism for More Reliable Search Engine Results, International Journal of Information Technology and Computer Science (IJITCS), Vol.7 (03), pp. 12–17 (2015).
- [4] Jonna Holland (2010), the role of mobile marketing communications in media strategy, Innovative Marketing, Volume 6, Issue 2, 2010.

- [5] Barathi, J.J.; Kavitha, G.; Imran, M.M., Building a Mobile Personalized Marketing system using multi-dimensional data, 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), pp. 133–137 (2015).
- [6] Leppäniemi, M., Sinisalo, J., and Karjaluo, H. (2006). A review of mobile marketing research, *International Journal of Mobile Marketing*, Vol. 1 (1), pp. 30–40.
- [7] Bruner II, G., and Kumar, A. (2005). Explaining consumer acceptance of handled internet devices, *Journal of Business Research*, Vol. 58 (5), pp. 553–558.
- [8] Pihlström, M., and Brush, G. (2008). Comparing the perceived value of information and entertainment mobile services, *Psychology & Marketing*, Vol. 25 (8), pp. 732–755.
- [9] Koren Y. (2010). Factor in the neighbors: Scalable and accurate collaborative filtering. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, Vol.4 (1):1, 2010.
- [10] Costanzo, A.; Faro, A., A fuzzy mobile recommender system: JQMobile vs FlashBuilder implementations. 2012 IEEE 3rd International Conference on Software Engineering and Service Science (ICSESS), pp. 513–518 (2012).
- [11] Ahn H.J. (2008). A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem *Information Sciences*, 178, pp. 37–51, 2008.
- [12] Razeq M.A., Frasson C., Kaltenbach M, Pyramid collaborative filtering technique for an intelligent autonomous guide agent, *International Journal of Intelligent Systems*, Volume 22 Issue 10, pp. 1065–1154, 2007.
- [13] Rafiul Hassan M., Maruf Hossain M., James Bailey and Kotagiri Ramamohanarao (2008), Improving k-Nearest Neighbour Classification with Distance Functions Based on Receiver Operating Characteristics, W. Daelemans et al. (Eds.): *ECML PKDD 2008, Part I*, Springer-Verlag Berlin Heidelberg, LNAI 5211, pp. 489–504, 2008.
- [14] Ki Joon Kim, “Can smartphones be specialists? Effects of specialization in mobile advertising Conference” *Journal of Telematics and Informatics* (2014).
- [15] Drossos, D., G.M. Giaglis, G. Lekakos. An Empirical Assessment of Factors that Influence the Effectiveness of SMS Advertising. *Proceedings of the 40th Hawaii International Conference on System Sciences*, 2007.
- [16] Leppäniemi M., J. Sinisalo, H. Karjaluo. A Review of Mobile Marketing Research. *International Journal of Mobile Marketing*, Volume 1, pp. 2–11 (2006).
- [17] ABI Research. Mobile Handset Demand Fuels 336.5 Million Shipments, 2010–29 January, <http://www.abiresearch.com/press/1593-Mobile+Handset+Demand+Fuels+336.5+Million+Shipments> [Last accessed at 29 December 2010].
- [18] Sinisalo, J., H. Karjaluo. Mobile Customer Relationship Management: a Communication Perspective. *International Journal of Electronic Customer Relationship Management*, Volume 3. – pp. 242–257, 2007.
- [19] Yang, K.C.C. Exploring Factors Affecting Consumer Intention to Use Mobile Advertising in Taiwan. *Journal of International Consumer Marketing*, Volume 1, pp. 33–41, 2007.
- [20] Yunos, H.M., J.Z. Gao, S. Shim. Wireless Advertising’s Challenges and Opportunities. *IEEE Computer*, Volume 36, pp. 30–37, 2003.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Barriers of knowledge acquisition in telecommunications companies in Saudi Arabia: An exploratory study on Etihad Etisalat Mobily

Mohammad Moeed Al-Harithy

King Abdul-Aziz University, Jeddah, Saudi Arabia

ABSTRACT: This research highlights accurately scientifically on knowledge acquisition obstacles in communication companies and especially on Saudi Mobily communication union company, then the study aims to identify important obstacles that faced officials, it determines their acquiring necessary knowledge to practice their business and to provide the necessary recommendations for the company to limit these obstacles. These obstacles divided in to three pivots, technological, human and organizational obstacles, whereof every pivot has a number of connected obstacles. Thus, the researcher used descriptive curriculum to treat this research problem and he used the questionnaire in this study for its suitability to such this explorative studies, for examined number of 80 word.

The researcher concludes via statistical analysis that human side is the important obstacle of knowledge acquisition generally in the first grade, he highlights this greatly in the first pivot (frightened of losing positions by the managers, lead the officials to non-participation in knowledge which they are distinct in other than others). It followed by (there is no helping by experts for the officials) then (an existence of feeling of functional security in the company).

Keywords: knowledge, knowledge management, knowledge acquisition, knowledge obstacles

1 INTRODUCTION

Perseverance endeavor for acquiring knowledge is not an indicator on leadership in knowledge community unless it built on new knowledge creation, but it evidences on pursuing obtaining what is generate from knowledge. Consequently, knowledge generating process is encompassing in the acquiring process form two sides when it is generate by others and while building on it to generate new knowledge.

A researcher mentioned [1], we are in era of knowledge which known origin of main economic assets in comparison towards traditional economic assets (workers, land, capital), this identification indicates to knowledge from two economic sides, which we see it is the basic economic origin which affecting on the remaining traditional factors in the production. How the capital grows and gets without knowledge? How land invests without knowledge? Moreover, how workers are train without knowledge?

The new economic is built on knowledge begins in changing traditional factors equation pillars. Hence, we would like to write about knowledge management that means all activities and processes that are performed on knowledge inside and outside the organization for maximum benefit of

this knowledge, knowledge acquisition is considered within these operations that we inquire what is hinder our acquisition of knowledge in our organization? What is prevent us to convey to its new era? What is concept of acquiring it? The researchers mentioned the concept of acquiring knowledge [2] which means what is included in all processes which are done by the organization to obtain knowledge from its different sources, whether it is obvious or figurative, it includes abstract, capture, purchase, research and exploring inside or outside the organization.

In our perspective, we think that generation and retrieval processes, research and exploring process are self-independent operations, which do not interfere within acquisition process but supporting of it. Hence, [3] identified knowledge acquisition “obtaining of the organization to the knowledge that are acquired by its officials, gathering and collection of it, creation of forms for it, assurance of its accurate for using of it in knowledge management programs and knowledge engineering programs”.

We think that concerning what is mentioned here previously that, acquisition process doesn't restrict on the organization officials, but from different directions whether from other organization or individual or research and consultant centers or

governmental or endowment agencies and other than that. As we mentioned previously, researchers' diversities majors (or specializations) result into diversity opinions and we think we can identify in acquiring knowledge from its resources and conveying of it to another receptacle. Whether it is obvious or figurative source for benefit and development of it for the sake of the organization as it considered non-palpable capital that can invested.

The benefit of acquiring knowledge is considered as seen by researchers [4] it contributes in change process, the internal education allows in development of the company internal abilities, so the external education is required to develop wider knowledge base, to for acquiring knowledge from its internal and external sources.

The companies are in need of continuously diverse and new external knowledge; therefore, the organization require gaining knowledge to avoid knowledge Obsolescence to keep on change speed. Then we see that knowledge acquisition has many non-countable benefits. such as knowledge acquisition create for the organization building of base for launch to creation and production of new knowledge, opening of new horizon for the organizations for leaderships in its sectors and giving it great competitive privileges, development of capital and increases the profit and it helps in investment of knowledge economically for attainment of its goals.

As for knowledge acquisition requirement [5] spoken about what we have to consider in our point of view when we want to think in performing knowledge acquisition project? Set of this in several points in a purpose of that layout results should be benefit for the final beneficiaries. Therefore, the officials in the organization have to benefit from the acquired knowledge and applying it at the work. should be conveyed to those who are in need of it without random arrangement to avoid adding additional cost on the organization, the project therefore shall be provided in high competency, whereon all the used resources should be available anytime for reviewing, and the project should not in trouble to lose time of experts.

As for the related knowledge acquisition barriers and obstacles that hinder acquiring of it. We referred that successful knowledge management contributes in generating new knowledge for the organization and it supports its competitive side in its anticipated sector. So any obstacles for knowledge management are considered as obstacles for acquisition process whereon [6] indicated in their speech about critical successful factors of knowledge management, they divided it in internal technological, financial, human resources, procedures and frames and cultural factors for infrastructure, strategies, leadership and knowledge management operations.

As for the related external factors, they divided to high effect factors such as (globalization, technology, education, sociology, politic, economic and laws) and intermediate effect factors such as companies, aliens, importers and referral comparisons with similar organizations. The researcher [7] spoken about knowledge acquisition obstacles he mentioned that there are two important factors in disabling external knowledge acquisition, the first run-back to knowledge and information source regarding effectivity, competency and operations in knowledge management. The second obstacle concerning to intellectual ability on effectivity, quality and content management. The researcher [8] also spoken about knowledge acquisition in different companies. He referred that the acquisition obstacles may differ from manufacturer to another by nature of the industry and its characteristics. He referred that there are three obstacles may defect knowledge acquisition. Such as non-existence of allowance system and incentives to and this enters in social side with the officials, existence of bureaucratic system to control the annual incentives and allowance system and others, therefore existence severe restricted financial system in the project that doesn't all or give any costs for acquisition process and knowledge transferring.

The researcher [9] indicated in his speech to effective factors of knowledge. He means the factors that encourage knowledge acquisition process and generation of new knowledges where he thinks that lack of availability to it may change it to obstacles. Such as no existence of development and research centers, lack of qualified human resources, no existence of administrative and financial and incentives, unique educational support as well as no existence of the social, economic and political settlement.

We can conclude some of knowledge acquisition obstacles whether it is internal source or external. Such as culture inside and outside the organization, in language, communication process, non-good availability access system, infrastructure, organizational structure, the followed policy in the organization, lack of support by higher management incentives no existence of development and learning support, fearfulness of missing distinctive in knowledge in case of participation with others and many several of obstacles. I divided obstacles in this study into technological, human and organizational obstacles.

2 THE PREVIOUS STUDIES

Several studies spoken directly or indirectly about the barriers of knowledge acquisition in the organizations. We are going to talk about these studies

to mention its objectives and the most important researches as A descriptive study titled by (**Technological development and its role in the management of knowledge in the Business Organizations**) [3]. A case of the general directorate for the communication establishment Algeria the study tends to answer the following main question. What extend had the organization of works applied the management of knowledge perfectly? How it can benefit from the technological development in this field? What does the management of knowledge mean? What are its factors? What are the barriers and challenges that faces it in the organizations of work? What are the factors that assist in the understanding and activating of the management of knowledge? Moreover, what is the role of technological development as one of these factors? The researcher used the analytical descriptive method. For the theoretical side.

The most important results of the study is that the organization depends on the main principles of knowledge and apply it in a rational way. (Acquiring, storing, distributing, and using of the knowledge). The method of research and development to acquire knowledge together with some staff specialized in data engineering within the human resources is a good support or earning knowledge.

In a study titled by (**Requirement of sharing knowledge and the hindrances that confront its application in the Jordanian telecommunication companies**) [10]. One of the objectives of the study to know the degree of the application of sharing knowledge in the telecommunication companies and to identify the most important factors that prevent the sharing of knowledge.

The results showed the impact of the availability requirements and elements of knowledge sharing combined (personnel training, team work, places of store knowledge, collaborative environment, obstacles of knowledge sharing). Moreover, the impact of teamwork and training as requirements individually to share knowledge on the application and practice of knowledge sharing in the Jordanian telecommunications companies, do not affect places of knowledge and collaborative storage environment as requirements individually to share knowledge on the application and practice of sharing knowledge in the Jordanian telecommunications companies. Finally, the knowledge sharing obstacles affect the application and practice of sharing knowledge in the Jordanian telecommunications companies.

Another study titled by (**A case study: Barriers preventing the capture of tacit Knowledge in small manufacturing companies**) [11]. The study done in the state of Michigan USA. The study aimed to find out the reasons of the failure of the managers and supervisors in the small manufacturing compa-

nies, to capture tacit knowledge from the employees. The theoretical base of the study depended on the four principles for forming knowledge, as said by (Nonaka and Takeuchi's). One of the targets of the study is to provide information that helps the small manufacturing companies to reduce the loss of tacit knowledge at the time when the employees leave the work. The questions were like: What is the awareness about the importance of acquiring and keeping the knowledge?, What is knowledge management policies, methods and procedures that can be designed for small businesses to support the capture and retention of tacit knowledge of employees?

The study concluded there were a number of obstacles that stand in the way of the ability to identify and capture knowledge. Including the lack of awareness by senior management regarding the importance of capturing tacit knowledge, lack of awareness by the supervisors on the first line with respect to the importance of capturing tacit knowledge, extreme lack of policies and procedures support for the management of tacit knowledge and lack of curriculum to manage tacit knowledge designed for small businesses.

Another study titled by (**Barrier to knowledge Acquisition Transfer and management in Regional knowledge Economy Development**) [8]. The study aimed to study two important factors to have a positive impact in the acquisition and transfer of knowledge and supporting learning and innovation in the knowledge economy of the former organizations. The first is the sustainability of the human capital of the workforce in knowledge and the second is sustainability of exchange of knowledge transfer between organizations. The study focused on the acquisition of knowledge, production, transfer and management by individuals, organizations and between organizations and study was in Pennsylvania in the United States between 2006–2011.

The study results indicated that the obstacles to the development of knowledge economy related to the development of human capital include employment is not appropriate for learners who misses the high technical skills and who misses teamwork skills as well as cultural barriers that do not support continuous learning and not learning culture. Moreover, continuous training of others and not to accept the cultural diversity of the workforce that appears in the inability of some small and medium sized companies to manage this diversity. The barriers of knowledge transfer between organizations referred to the lack of information and communications technology that links between similar sectors. Another problem in knowledge flow between universities and industry represented in the absence of a social networking case of organizational turmoil, the variation in

power asymmetries in understanding organization goals, by the last differing institutional cultures like different rewards system. All these obstacles impede the acquisition of knowledge transfer.

In a study titled (**Barriers and Facilitators to Knowledge Capture and Transfer in Project-Based Firms**) [12]. The study aimed to increase understanding of the policies of knowledge management in organizations based on projects of human resources management perspective.

The results indicated the different policies of companies in knowledge capture and transfer. Some companies' depends on Incentives to control their employees socially. Another depends on structured incentives bureaucratically and finally found that the one airline has relied on a strict financial system work as a barrier of knowledge management and their Processes because of no funds to capture the knowledge and activities transferred.

3 THE RESEARCH PROBLEM

As for the long experience of the researcher in the field of communication comes the opinion of this research. One of the main difficulties is that, the foreign experts never support the local Saudi employee, by providing them with knowledge and skills to enable them to work on modern systems. That Situation decrease knowledge sharing. This talking supported by study [13], which indicated that in high-tech environments, might reduce staff participation and sharing knowledge with each other and telecommunications companies considered as high-technology environments.

The main question is what are the barriers of acquiring knowledge in Saudi communication companies?

The secondary questions are:

- a. Can we consider the organizational factors as barriers? (Training, Mastering English language, Establishment of knowledge administration, communication inside the organization, incentives, development in the levels of Education).
- b. Can the human factors work as barriers? (Support of experts, fear of sharing knowledge that may lead to the loss of post or position)
- c. Can the lack of technology be a barrier? (Networks, Integrated information systems)

4 RESEARCH METHODOLOGY

The communications sector considered the fastest growth and development sector in modern era. This study aims to know knowledge acquisition obstacles in communication sector and particularly in Mobily

Communication Co. from the official's point of view in the company. The researcher used questionnaire as a tool for the study. The concentration was on three main factors (Technological, Human and Organization Factors), a number of variants is merged with every factor. These variants are abstracted from previous studies such as [14], [15], [6] & [16] where 80 questionnaires were distributed 72 were returned and 4 were dismissed for incomplete. The study has been done at the regional management as well as the main purchase official office and purchase office at Al-Andalus Dist. Branch in a number of 45 official, through the period from 06/12/2015–20/12/2015.

The questionnaire final copy contained the following parts:

First Part: It includes initial data about the Research sample represented in: Position, scientific qualification and experiences years.

Second Part: It consists of the study tools, which relates to knowledge acquisition obstacles in communication companies, it includes (11), statement that have divided in three pivots:

First Pivot: Organizational obstacles which includes (6) statement.

Second Pivot: Human obstacles which includes (3) statement.

Third Pivot: Technological obstacles which includes (2) statement.

The researcher used Likert scale (Extremely Non-Acceptance—Non Acceptance—Neutral—Accepted—Extremely Accepted) for determination of knowledge acquisition obstacles in communication companies.

5 STATISTICAL ANALYSIS RESULTS

Standard deviations and arithmetic medians calculated for answering the study questions (technological, human and organizational obstacles), schedule no. (1) Indicates the arithmetic medians for these pivots.

From the above schedule the axis of human obstacles came in first place in terms of the approval of the highest arithmetic average was (2.86), followed in second place regulatory obstacles an arithmetic mean was (2.28), and ranked the last obstacles technical and an arithmetic mean was (2.16).

5.1 Organizational obstacles

Do organizational factors unavailability from (development of officials' educational level, education incentives, obviousness structures and communications inside the organization, knowledge management quality, English Language and training) hinder knowledge acquisition in Mobily Union Co. from officials' point of view?

It began obvious in schedule (2) that the organizational obstacles elements from Mobily Communication Co. workers point of view. Regarding arithmetic medians registered in arithmetic medians of the questionnaire degree of "Acceptance", the arithmetic medians of the elements registered between (3, 49–3, 97). Where the highest degree of acceptance from sample personnel point of view of elements (the company commit in training on new technology, whereon it registered an arithmetic median of 3, 97). The other elements that obtained less medians with some of contrast in the acceptance degree. It includes: "there is knowledge management in the company" (M = 3, 90), "the company commit to create language preparation for the official such as (English Language)" (M = 3, 68), "there are incentives for development and learning in the company" (M = 3,66), "the officials' educational level is developed" (M = 3,65).

The less acceptance degree from point degree of the sample personnel for the elements (there are obviousness in organizational structures and communication directions were between the different managements inside the company, whereon it registered the arithmetic medians of 3.49).

5.2 Human obstacles

Do human factors hinder from (experts support to the officials, lack of functional security feeling and worrying of losing the position by knowledge participation reasons) knowledge acquisition process in Mobily Communication Union Co. from officials' point of view?

It began obvious in schedule (3) that the human obstacles elements from Mobily Communication Co. workers point of view. Regarding arithmetic medians registered in arithmetic medians of the questionnaire degree of "Acceptance". The arithmetic medians of the elements registered between (3,40–3,69) whereof the highest degree of acceptance from sample personnel point of view of elements (fearfulness of losing positions by managers lead to non-participation in the knowledge that they are distinct in comparison than others) and it gained arithmetic medians of 3,69. In addition, this supported by research [13]. Saying that in the higher technologies environments, officials may decrease in participation and dividing knowledge with each other's. And the companies of communication is considered within higher technological environment. As for other elements that obtained

Schedule No. (1) Standard deviations and arithmetic medians and arrangement for knowledge acquisition obstacles in communication companies.

Pivot	Median	Deviation	Arrangement	Acceptance degree
Organizational obstacles	2.28	0.897	2	Non Accepted
Human obstacles	2.86	0.773	1	Neutral
Technological obstacles	2.16	0.904	3	Non Accepted
Questionnaire as whole	2.42	0.781		Non Accepted

Schedule No. (2) Standard deviations and arithmetic medians and arrangement for organizational obstacles elements for knowledge acquisition.

Seq.	Elements	Arithmetic median	Standard deviations	Arrangement	Acceptance degree
1	Company oblige training on new technology	3.97	0.992	1	Accepted
3	There is knowledge management in company	3.90	0.949	2	Accepted
2	Company oblige in language preparation for official such as (English Language)	3.68	1.085	3	Accepted
5	There is incentives for development and learning for officials	3.66	1.128	4	Accepted
6	Education level development is done for officials	3.65	1.255	5	Accepted
4	There is obviousness in organizational structures and communication trends between the different managements inside the company	3.49	1.191	6	Accepted
	Organizational obstacles as whole	2.28	0.897		Non Accepted

Schedule No. (3) Standard deviations and arithmetic medians and arrangement for human obstacles elements for knowledge acquisition.

Seq.	Elements	Arithmetic median	Standard deviations	Arrangement	Acceptance degree
8	Fearfulness of losing positions by managers lead to non-participation in the knowledge that they are distinct in comparison than others.	3.69	1.175	1	Accepted
7	There is support and help of experts for the officials	3.66	1.087	2	Accepted
9	There is feeling of functional security in the company	3.44	1.1250	3	Accepted
Human obstacles		2.86	0.773	Neutral	

Schedule No. (4) 10 Standard deviations and arithmetic medians and arrangement for technical obstacles elements for knowledge.

Seq.	Elements	Arithmetic median	Standard deviations	Arrangement	Acceptance degree
10	Company provides integrated electronic networks for facilitating access to knowledge and information sources	3.90	0.964	1	Accepted
11	Company has integrated information systems to support its different sectors	3.78	1.005	2	Accepted
Technical obstacles		2.16	0.904	Non Accepted	

less medians with some contrast in acceptance degree. It includes “there is support and help of experts for the officials” (M = 3, 66), it was a less acceptance degree from point of view of the sample personnel for the elements (there is feeling of functional security in the company and whereon it registered the arithmetic medians of 3.44).

5.3 Technological obstacles

Do technological factors unavailability hinder from (integrated information systems and networks) knowledge acquisition process in Mobily Communication Union Co. from officials’ point of view?

It began obvious in schedule no. (4). that the technological obstacles elements from Mobily Communication Co. workers point of view concerning arithmetic medians registered in arithmetic medians of the questionnaire degree of “Acceptance”. The arithmetic medians of the elements registered between (3,90–3,78) whereof the highest degree of acceptance from sample personnel point of view of elements (the company provides integrated electronic networks for facilitating access to knowledge and information sources) and it gained arithmetic medians of 3,90.

The less acceptance degree from point degree of the sample personnel for the element (there are obviousness in organizational structures and com-

munication directions were between the different managements inside the company, whereon it registered the arithmetic medians of 3.49). The company provides integrated electronic networks for facilitating access to knowledge and information resources for supporting its different sectors and it gained arithmetic medians of 3,78.

6 RECOMMENDATION AND CONCLUSIONS

It began obvious via statistical analytic results that most important obstacles for knowledge acquisition generally is the human side. In the first stage, this highlights greatly in the first pivot (fearfulness of losing positions by managers lead to non-participation in the knowledge that they are distinct in comparison than others), then followed by (there is support and help of experts for the officials). Then (an existence of feeling of functional security in the company).

We indicate for need of disseminating knowledge participation culture between the officials inside organization environment and the need of using functional circulation for no domination of positions. It is necessary to encourage financially and morally those who having knowledge from the experts and managers and others for participation of knowledge and transferring of it for others within the organi-

zation. There is also necessity to let officials feel of functional security, this may be through entering in long contracts with those who are distinct.

It is without doubting that technological and organizational factors have an important side as factors which support knowledge, lack of availability hinder completely knowledge management, so it is notable Mobily Communication Union Co. are obviously interested in as result of this study.

Finally, more research a question needs. Do we reach with our organizations in Meddle East to knowledge awareness that qualify us change to knowledge community?

REFERENCES

- [1] Montoro, Mario. (2004). The Knowledge Identification Problem: Scope and Consequences in Network Society. Retrieved February 06, 2016 from: ftp://ftp.cordis.europa.eu/pub/ist/docs/perez_montero_the_knowledge_identification_problem_workshop_brussels_7_and_8_june_2004.pdf.
- [2] الحارثي، سعد. (2009، 4-1 نوفمبر). أنموذج مقترح لتطبيق إدارة المعرفة في القطاع الحكومي في المملكة العربية السعودية. المؤتمر الدولي للتنمية الإدارية، معهد الإدارة العامة، الرياض.
- [3] نوي، طه حسين. (2011). التطور التكنولوجي ودوره في تفعيل إدارة المعرفة بمنظمة الأعمال: حالة المديرية العامة لمؤسسة اتصالات الجزائر. رسالة دكتوراه، كلية العلوم الاقتصادية والعلوم التجارية وعلوم التسيير، جامعة الجزائر، الجزائر.
- [4] Saez, Pedro & Lopez, Jose & Castro, gregorio & Gonzalz, Jorge. (2010). External knowledge acquisition processes in knowledge-intensive clusters. *Journal of Knowledge Management*, Vol. 14 Iss: 5, pp. 690–707.
- [5] Milton, r. (2007). Knowledge acquisition in practice: a step-by-step guide. Retrieved February 13, 2016 from <http://books.google.com.sa/books?id=wHf67vdK-QUC&pg=PR4&lpg=PR4&dq=Knowledge+Acquisition+in+Practice:+A+step+by+step+guide,+Milton,+Springer-Verlag&source=bl&ots=uccOcSk3S5&sig=6SBggg8tQnUEHNOfaHcMZs23J1A&hl=ar&sa=X&ei=GXVoVK7cFcbnygP564DoAw&ved=0CD0Q6AEwBQ#v=onepage&q=Knowledge%20Acquisition%20in%20Practice%3A%20A%20step%20by%20step%20guide%2C%20Milton%2C%20Springer-Verlag&f=false>.
- [6] Sheen, Margaret. (1992). Barriers to scientific and technical knowledge acquisition in industrial R&D. *R&D Management Journal*.vol 22. issue 2. pp. 135–143.
- [7] Trauth, Eileen M. (2012). Barriers to Knowledge Acquisition, Transfer and Management in Regional Knowledge Economy Development. 45th Hawaii International Conference on System Sciences. Hawaii. USA.
- [8] عليان، ربحي. (2008). إدارة المعرفة. (ط1). عمان: دار صفاء للنشر والتوزيع.
- [9] الطاهر، اسمهان & منصور، إبراهيم. (د.ت). متطلبات مشاركة المعرفة والمعوقات التي تواجه تطبيقها في شركة الاتصالات الأردنية. <http://eco.asu.edu.jo/ecofaculty/wp-content/uploads/2011/04/20.doc>.
- [10] Sherwood, William Roy. (2013). *A Case Study: Barriers Preventing The Capture Of Tacit Knowledge In Small Manufacturing Companies*. Ph.D. thesis in Business Administration Information Systems, Baker College, Flint, Michigan.
- [11] Hall, Jeremy & Sapsed, Jonathan & Williams, Kelly. (2000,28–31 August). *Barriers and Facilitators to Knowledge Capture and Transfer in Project-Based Firms*. 4th international conferences on technology polices and innovation. Curitiba.
- [12] Smoyer, Eric W. (2009). *Identifying Knowledge Sharing Relationships In The Telecommunication Industry*. Ph.D. thesis of management in organizational leadership with a specialization in information systems and technology. university of phoenix. USA.
- [13] Oliva, Fabio. (2014). Knowledge management barriers, practices and maturity model. *Journal of knowledge management*, vol 18, no 6, pp. 1053–1074.
- [14] Frost, Alan. (2014). *A Synthesis of Knowledge Management Failure Factors*. Retrieved February 11, 2016 from: [file:///C:/Users/only%20me/Downloads/www.knowledge-management-tools.net_a_synthesis_of_knowledge_management_failure_factors%20\(1\).pdf](file:///C:/Users/only%20me/Downloads/www.knowledge-management-tools.net_a_synthesis_of_knowledge_management_failure_factors%20(1).pdf).
- [15] Ajmal, Mian & Helo, Petri & Kekale, Tauno. (2010). Critical factors for knowledge management in project business. *Knowledge Management Journal*. vol.14. No.1, pp. 156–168.
- [16] Sedighi, Mohammad Bashir & Zand, Fardad. (2012, 21–23 November). *Knowledge management: Review of the Critical Success Factors and development of a conceptual classification model*. ICT and Knowledge Engineering 10th International Conference. Bangkok.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

The effect of using computerized educational software based on interactive video in developing some of using computer skills for the preparatory year students at Albaha University, Saudi Arabia

Rajeh Saad Abdullah AlSelouly
Albaha University, Saudi Arabia

ABSTRACT: This article aimed to investigate the effect of using computerized educational software based on interactive video in developing some of using computer skills for preparatory year students at Albaha University. The researcher has identified computer skills through a list of skills that were judged by professionals in education and computer technologies, and then designed educational software prepared by the researcher using interactive video and operated by several computer programs. Finally, achievement test and note card were prepared as measure tools of the study. This article contributes to education technology departments by providing software for developing collection and practical skills in less time than it takes in normal ways of teaching using computer subject, and contribute to utilize interactive videos in teaching IT skills in preparatory years, as well as provide curriculum Education technology planners with an technological computerized educational content that achieves important objectives in the field for student benefit. Finally software contributes to identify using computer skills that will allow interactive software to communicate with students at any time.

The researcher used the experimental method with quasi-experimental designed, which comprises of two groups that were selected randomly, the experimental group (which used educational interactive video software) and the control group (which used traditional method). The study sample consisted of 48 students, evenly divided into the study groups (experimental and control) in each group there are (24 students). The researcher applied a tribal pretest on the two groups to ensure the homogeneity of the two groups in the study subject. After application of the experiment was completed, a posttest was taken by the study groups to compare the results. After experiment period which lasted for six-week results indicated statistical significance differences at the level of significance (0.01) between average scores of the experimental group and control group in the posttest, this means that the program has significant effect on students performance of the experimental group. The researcher also proposes a set of recommendations which may contribute to using of interactive video to teach any skills that are of benefit to student.

Keywords: education technology, interactive video, education, computer skills, Albaha University

1 INTRODUCTION

Current era is featured by educational technological progress and production of many technological innovations that have played an important role in increasing educational process and efficient development. This development has led to emergence of education systems and increasing trend to utilize them in educational process, including: a single-learning, personal education system, distance learning, computer-aided instruction, based on multimedia learning, and other systems, tools and instruments that used in educational process.

It is reported [1] that e-learning is of modern method in and education techniques that provides the latest techniques of hardware and software in the education process, ranging from use of elec-

tronic presentation means which are used to teach lessons in traditional classrooms and use of multimedia in class room education and self-learning processes and ending with building smart schools and virtual classrooms that allow students to attend and interact with lectures and seminars that are held in other countries through the Internet and video conferencing technologies.

It is reported also [2] that use of interactive video teaching has characteristics of both video and assistant computer, as displayed audiovisual information by video represent reality and can offer expertise and skills that computers cannot perform alone. Computers provide an interactive environment which represented the ability of learners to control self-speed, the path followed through the program and information progress as well as

computer's ability to provide an immediate return to learner response. This interaction provided by computers in interactive video programs is the part that missed in linear videos, which present programs or movies in linear manner that does not allow control and choice. Programs here are integrated unit that are displays in logical order starts from beginning until the end, on the other hand we note that computers provide control and choice property on programs. From the foregoing; it can be said that use of educational software based on interactive video teaching has characteristics and advantages of both video and assistant computer for education; it also allows learners to control the display and sequence information through an active interaction between learners and educational material provided through computers.

Base on importance of developing of student skills in computers, educational experiment was done for educational software based on interactive video between students and computers in education and designing of educational software to measure effect of computerized educational software use that base on interactive video to develop some of using computer skills at the preparatory year students of Albaha University courtyard, where basic computer skills were identified through a survey of teaching staff views in Computer Department who are in charge of preparatory year of Albaha University, then calculating of questionnaire reliability and validity to be agreed upon. In this paper the researcher will review the research question, its significance, its methodology, and its population and sample, as well as review literature and digital results. Finally, presents the study results and its recommendations and proposals.

2 LITERATURE REVIEW

Study [3] entitled "Effectiveness of mini-teaching proposed program based interactive video technology in developing of teaching executive skills to students at the Education College, Jazan University," which aimed to: prepare a list of executive skills to be met by student teachers to teach in classrooms which is derived from opinions of experts and specialists in curriculum and teaching methods, and building of a mini-teaching proposed program based interactive video technology and the study its effectiveness in developing skills of lesson implementation the classroom. Experimental method with a quasi-experimental design is used, and the sample is grade eighth scientific department students who are eligible for application on practical education program at Education College, Jazan University. The study found, statistical significance differences at level of (0.01)

between averages and pre and post tribal application degrees for card note are ranked in favor of pre-application, which indicates the effectiveness of proposed program in developing implementation skills of lessons into classrooms.

This study is associated with the current study, in respect of benefiting from a list of skills.

Study [4] entitled "The effectiveness of programs which using interactive video" in teaching of comprehensive libraries subject and lasting of learning effect and acquisition of dealing skills with learning resources for Education Technology students at the Education Faculty, South Valley University, which aimed to investigate effectiveness of a program that uses interactive video to teach comprehensive libraries subject and lasting of learning effect and acquisition of dealing skills with learning resources for Education Technology students at the Education Faculty, South Valley University. The study sample consisted of (80) students (males and females), who were chosen purposely, and has concluded its findings with presence of statistical significance differences at level of (0.05) between the two groups in post-test tests for the benefit of the experimental group, which studied through using interactive video technology. The study also found presence of statistical significance differences at level of (0.05) between the two groups in lasting of learning effect for benefit of the experimental group that studied using interactive video technology, as well as of presence of statistical significance differences at level of (0.05) between the two groups in respect to practical skills of dealing with learning resources for benefit of the experimental group that studied using interactive video technology.

This study is associated with the current study, regarding using of a program that based on interactive video.

Study [5] entitled "The effect of using interactive video on the developing scientific trends of the fifth grade primary school students in Jordan," This study aimed to investigate effect of using interactive video in developing scientific trends for fifth grade students. The study sample consisted of (52) students, who were distributed into two groups: an experimental group consisted of 27 students, who studied using interactive video technique, and the control group which consisted of 25 students who studied in traditional manner.

The researchers set up the educational situations, which included (60) computer chip, validity and reliability of which has been confirmed to answer the study according to scientific trends scale before the study (pre-test), and then after it to find the effect of interactive video teaching way. One-way analysis of variance (ANOVA) at significance level of ($\alpha = 0.05$) is used. The study

results have revealed the positive effect of interactive video where the experimental group members prove superior to the control group in the scientific trends. This study is associated with the current study, in respect of using a program based on interactive video.

Study [6] entitled “Designing and producing educational software equipped with interactive video technology, on some of the basic skills of football game for second grade students who are studying basic education”. The study aimed to design and produce educational software for some football skills under investigation which is equipped with interactive video technology through educational program and to find its effect. The researcher has used the experimental method with a sample search of (40) students who are divided into two groups (experimental and control). One of most important results was that; software equipped with interactive video technology has contributed in a positive way to improve technical and skill performance level and cognitive achievement for skills such as running with ball for members of experimental group. This study is associated with the current study, regarding design of a program which base on interactive video.

Study [7] entitled “Evaluating interactive video users perceptions of self access language learning with multimedia movies”, which aimed to identify the extent of using possibility of interactive video to enrich learners experiences and knowledge that will help to learn some sports activities. The Researcher has used experimental method, and sample was used as two groups one was experimental which educated through interactive video and control group was taught in a traditional way. One of the most important results that, learners through interactive video have achieved results higher in learning some sports activities because they enrich their experience, further programs are fun and interesting. This study is associated with the current study, in respect of using a program based on interactive video.

Study [8] entitled “Does project L.I.E. Case modality impact on critical thinking in PBL groups. A paper presented at the American Educational Research Association”, which aimed to investigate the impact of education project with interactive video (L.I.V.E) on critical thinking in education—groups based on a problem. The researchers have developed and tested a new model to provide—based—problem learning cases which is a CD-ROM (a computer program on the internet) that uses interactive video project known as (L.I.V.E). The study strived to answer the question, “Is students critical thinking varies during learning discussions according to discussions way”. 128 students has been involved in this study, who

were distributed into three groups, a face-to-face group which used a written text on papers, a group face-to-face which used videos, and a virtual group which used digital video. The study results was that, the virtual group merged in critical thinking more than other two groups, which the researchers have interpreted as increasing of individual responsibility that discussion required on the internet. This study is associated with the current study, in respect of using a program based on interactive video.

3 PROBLEM STATEMENT

Literature review conducted on interactive video reflects the fact that students who studied using interactive video have achieved higher grades than students who studied using the traditional way in academic achievement of Information Technology (2) curriculum. Therefore it became necessary to design educational software based on video interactive that helps developing skills of using computer, as there are shortcomings in teaching skills of using computer in the preparatory year at Albaha University. As the researcher has observed used teaching method to teach the preparatory year students using computer skills, the idea of the study, that lecturer s explain steps of each element separately using software based on using of interactive video, was originated.

The study question states “what is the effect using computerized educational software based on interactive video in developing some of using computer skills for the preparatory year students at Albaha University?” which is the main question. From this main question the following sub-questions are emerge; first, “What is the necessary computer skills that you should develop in the preparatory year students at Albaha University?”, second, “What is the effect of using computerized educational software based on interactive video on developing of students achievement at the preparatory year at Albaha University?” And finally, “What is the effect of using computerized educational software based on interactive video on developing some of students using computer skills at the preparatory year at Albaha University?”

4 THE IMPORTANCE AND OBJECTIVE OF THIS WORK

This study contributes to providing education technology departments with a technological software for developing achievements and practical skills in less time than it takes through tradition ways of teaching using computer subject, and contribute to

the using of interactive video in teaching IT skills in the preparatory year, as well as providing curriculum Education technology planners with an technological computerized educational content that achieves important objectives in the field for student benefit. Technological programs may utilize in achieving a variety of educational goals, especially practical ones. Finally software contributes to identification of using computer skills that will allow interactive software to communicate with students at any time.

The main goal of this study is to identify the effect of using computerized educational software based on interactive video in developing some of using computer skills for the preparatory year students at Albaha University. This goal can be divided into the following objectives: first, the researcher recognizes necessary computer skills that must be developed in the preparatory year students at Albaha University, secondly, the researcher recognizes the effect of using computerized educational software based on interactive video which is proposed to increase the academic achievement of the students of preparatory year at Albaha University, finally, to the researcher recognizes the effect of using computerized educational software based on interactive video which is proposed to develop computer skills for students of the preparatory year at Albaha University.

5 THE METHODOLOGY OF SOLUTION

There are two assumptions for this study: first, “No statistical significance differences between the average scores of the two groups (experimental and control) in the post measurement of learning achievements of the preparatory year students of Albaha University after pre-performance is set”. The second assumption stipulates that “No statistical significance differences between the average scores of the two groups (experimental and control) in the post measurement of computer using skills among the preparatory year students of Albaha University after pre-performance is set.” The study base on using of experimental design as it contains two groups (experimental and control),

a list of using computer skills was used which function as a content analysis of the study. As well as achievement test to measure the cognitive side of skills and to ensure equality and homogeneity of both groups. Further a note card is used to measure the using computer skills for spreadsheets (Excel) program unit, and statistical program SPSS.

6 NUMERICAL RESULTS

Assumption of the first study states: “There are no statistical significance differences between the average scores of the experimental and control groups in post measurement of academic achievement of the preparatory year students at Albaha University.” To find out whether there is a statistical significance differences between average scores of the experimental and control groups in the post measurement of interactive video program in academic achievement of the preparatory year students at Albaha University, the researcher used the t-test for independent samples on experimental and control study sample after interactive video program, and the results were as illustrated in the following table:

The above table shows no statistical significance differences in grades of the control group and the experimental group in post measurement for academic achievement, at a level of significance of 0.01 or less. Therefore, it is clear that the use of proposed interactive video program by the researcher had an effective and positive impact on increasing of academic achievement for the preparatory year students of Albaha University. This is clearly indicates that first assumption was wrong, and acceptance of the alternative assumption, which is no statistical significance differences between average scores of the experimental and control groups in post measurement of academic achievement skills of the preparatory year students at Albaha University in favor of the experimental group students.

Assumption of the second study states: “There are no statistical significance differences between the average scores of the experimental and control groups in post measurement of computer using

Table 1. t-test for independent samples of differences between the views averages of experimental and control study sample in post measurement of achievement test grades.

Study axis	Sample	No.	Average	Standard deviation	Value (T)	Freedom degrees	Level of significance
Academic achievement	Control group	24	13.96	3.42	2.789	46	0.008** significance
	Experimental group	24	16.33	2.39			

**Significant differences at the level of 0.01 or less.

Table 2. t-test for independent samples of differences between the views averages of experimental and control study sample in post measurement of note card grades.

Study axis	Sample	No.	Average	Standard deviation	Value (T)	Freedom degrees	Level of significance
Computer using skills	Control group	24	27.38	7.04	6.33	46	0.00** significance
	Experimental group	24	39.00	5.60			

**Significance differences at the level of 0.01 or less.

skills level of the preparatory year students at Albaha University.” To find out whether there is a statistical significance differences between average scores of the experimental and control groups in the post measurement of computer using skills level of the preparatory year students at Albaha University, the researcher used the t-test for independent samples on experimental and control study sample after interactive video program, and the results were as illustrated in the following Table 2.

The previous table shows no statistical significance differences in grades of the control group and the experimental group in post measurement for computer using skills, at a level of significance of 0.01 or less. Therefore, it is clear that the use of proposed interactive video program by the researcher had an effective and positive impact on developing of computer using skills for the preparatory year students of Albaha University. This is clearly indicates that second assumption was wrong, and acceptance of the alternative assumption, with existence of statistical significance differences between average scores of the experimental and control groups in post measurement of academic achievement skills of the preparatory year students at Albaha University in favor of the experimental group students.

7 CONCLUSION AND RECOMMENDATIONS

In respect of first assumption of the study there was existence of statistical significance differences between the average scores of the experimental and control groups in post measurement in academic achievement skills of the preparatory year students of Albaha University in favor of experimental group students. With respect to second assumption of the study, there was existence of statistical significance differences between the average scores of the experimental and control groups in post measurement in computer using skills of the preparatory year students of Albaha University in favor of experimental group students.

In light of the study results, the following recommendations could be suggested: first, seek to develop the curriculums of the educational technology in order to urge learners to singular learning to cope with the growing large numbers of learning groups from time to time. Universities can be equipped with necessary devices for recording of laser discs, interactive laser televised discs, all tools and materials that interactive video programs needed in educational process. Finally, paying attention to advanced software, particularly in field of programming of computer languages for producing of programs that serve university education.

The study suggests a set of proposals: first, conducting a study on effectiveness of using interactive video in achieving of curriculum goals related to learning techniques. As well as further studies and researches to determine effectiveness of using interactive video in education at different educational stages, it can also carry out a study on impact of using software based on interactive video on academic achievement in different educational stages.

REFERENCES

- [1] Alkawas, Mahdi Mohammed, (2008), Toward practical model for producing university courses electronically, typical legal social science. Mansoura University, Faculty of Arts, Cairo: The fourth annual scientific symposium.
- [2] Al-Baghdadi, Mohammad Reza, (2002). Education and educational technology. Cairo: Dar Alfrk Alarabi.
- [3] Albarbari, Rafik Said, Isaac, and Hassan Abdullah, (2010). “Effectiveness of mini-proposed program for teaching based on interactive video technology to develop teaching practical skills for students of Education College at Jazan University,” Journal of Science Education, Volume 13, Issue 6.
- [4] Alsaid Sahar Muhammad, (2010). The effectiveness of a program by using interactive video in the teaching of the comprehensive libraries on lasting of learning and acquisition skills of dealing with learning resources for education technology students of

Faculty of Education in South Valley University. PhD thesis, Faculty of Education, University of Beni Suef, Beni Suef.

- [5] Alaqrarah, Ahmed Odeh, Alrvua, Muhammad Ahmad, al-Qaisi, and Taysir Khalil, (2007), "The impact of using interactive video on developing scientific trends for students of the fifth grade in Jordan", Journal of Qatar University for Educational Sciences, Issue No. 12.
- [6] Abdel Fattah, Mohamed Sobhi, (2005), "Designing and producing educational software equipped with interactive video technology on some of the basic football skills for second grade basic education students" Master research, Faculty of Physical Education, Tanta University.
- [7] Gardnare & Dave (2003). "Evaluating user interactive video users perceptions of self access language learning with multimedia movies", (China) open university United Kingdom.
- [8] Kamin carol, O Sullivan, Patricia Deterding, Robin (2002). Does project L.I.E. Case modality impact critical thinking in PBL groups. Paper presented at the American Educational Research Association, New Orleans.

The reality of tacit knowledge sources in educational training center in department of education of Bisha province: Saudi Arabia case study

Abdullah Mohammed Abdullah Alyateem

Department of Information Technology, King Abdulaziz University, Saudi Arabia

ABSTRACT: The objective of this article is to identify sources of tacit knowledge in Educational Training Center in Department of Education of Bisha province, and to identify the role of tacit knowledge sources in professional development in Educational Training Center in the Department of Education of Bisha province, in addition to the identification of the role of knowledge technologies in developing of tacit knowledge to promote professional performance in Educational Training Center in the Department of Education of Bisha province.

The question with this study is to identify the reality of these tacit knowledge sources in the field of education that offered through educational training centers, especially Educational Training Center in the Department of Education in Bisha province, this study seeks to answer the main question, which states, "What is the reality of tacit knowledge sources in Educational Training Center in the Department of Education in Bisha province?". The importance of this research come from determining the tacit knowledge sources that are currently available at the training center in addition to the identification of its role in the professional development in the field of education which gives us an idea and a base to launch towards developing educational training centers in the area of utilizing experience and tacit knowledge, depolarizing and effectively manage through utilizing private sources of tacit knowledge that serve the educational field.

In this research the researcher used the descriptive case study in order to answer the question and the research tools is interviews and questionnaires. The study finds that tacit knowledge sources have a role with rate of 70.00% in professional development in Educational Training Center in the Department of Education in Bisha province as well as the study finds that knowledge technologies play a role in developing of tacit knowledge to develop professional performance in Educational Training Center in Bisha province. Arithmetic average of that is equal to the (2.80), which mean knowledge technologies has a role of by 70.00% in development of tacit knowledge that develops professional performance in Educational Training Center in Bisha province.

Keywords: tacit knowledge, tacit knowledge sources, educational training, training center, knowledge technologies

1 INTRODUCTION

Human elements are the most important wealth and fundamental building block upon which the nation is built with its institutions and organizations, and any evolution of human elements lead to evolution of its organization, which leads in turn to contribute to development of a nation to which it belongs.

From here, Nations are very interested in issue of developing of human elements in their respective fields, the matter is independent of education stages, but headed towards the subject of in-service training to develop the human elements and raise efficiency work and increase the quality of the product, therefore each organization directs to establish a training center to serve its objectives and develops its staff.

The Ministry of Education was not immune to these trends, it has established the educational training centers in most of its educational departments, so that ministry employees whether they are teachers or school administrator or department heads or supervisors educators or administrative staff can be trained through these centers. Educational training centers like any other training centers concern and need knowledge in the field of education in order to be able to carry out its duties for which are created, including explicit and tacit knowledge. However talk about the explicit knowledge is of significance, but this part is clear as the fact that explicit knowledge is clear, available and possible to find and take advantage of it. The part of tacit knowledge required further clarification and discussion and stand on its reality in terms of resources and its role in training centers, and this is what we will consider in this study through

case study of educational training center in Bisha. The study question and questions will be reviewed, in addition to the study methodology and its population and samples, as well as a literature review related to the study topic, and then review the digital results and its interpretation. Finally, present the study results and its proposals and recommendations.

2 RELATED WORKS

Study [1] aimed to identify “the effectiveness of the educational training centers in developing of management performance for managers of primary schools in Jeddah, from their point of view.” The study aimed to identify the effectiveness extension and to reveal the significance of the differences between the mean estimate of managers of the educational training centers effectiveness in their administrative development that attribute to factors of age, academic qualifications, experience and training courses, as well as to determine the development of these centers requirements to achieve the administrative performance growth of the managers from their perspective. Analytical method is used and the results of the study suggest that the effectiveness of the educational training centers begin with planning needs of managerial training programs. It has been shown through the study that effectiveness and impact of training taking first place in respect to dimensions and then came implementation of training. Training has been assessed as moderate as well as administration follow-up and its support. In addition to lack of statistically significant differences in the estimation of directors due to the age, academic qualifications differences, years of experience and training courses. The study suggests required development that is needed and the necessity of providing expert supervisors for administrative programs, and instructors in educational administration and continuous communication between the training centers and school administrators with aim of development.

Study [2] aimed to build a proposed model for a system of educational training in the education management of for girls in Jeddah in the light of contemporary transformations. The researcher has applied two tools: a survey of some managers of educational supervision management departments in the education management of for girls in Jeddah, and questionnaire which a main tool. The study concluded with the suggested model for a system of educational training in the education management of girls in Jeddah. Features of this proposed model are parameters and techniques to identify training needs to perform many procedures. The most important of these procedures is allocation of planning unit and development supported by studies and researches unit shall perform many tasks

on the educational management level. As well as studying of economic transformations and technological and scientific developments that affect local systems; to determine the future training needs. At the level of educational training centers determine training needs for a trained specialized team on how to perform this task. This team is reported to training programs unit, which manages planning, design and production of training programs.

Study [3] aimed to identify trainers training needs in educational training centers in the holy Makkah, revealing of statistical differences importance regarding training needs degrees and fields that may be attributed to the educational qualification, experience, training fields, job title, and a full-time training. The researcher used descriptive methodology, and the study population consists of all instructors in educational training centers in Mecca totaling 183 instructors. Questionnaires were distributed among them in a manner of comprehensive inventory and 177 questionnaires recollected from distributed questionnaires who are individuals of the study population. The study results concluded to identification of the degree of training needs for instructors in general is high. The recruitment field of training management change is ranked as first, followed by the field of using of modern training techniques and using of strategic planning skills and then build a training package, then followed by the field of “identifying trainees needs” and then training programs design which ranked as the last one. Finally, follow-up, evaluation and measurement the impact of training which ranked as last.

Based on the findings of the study results, a number of recommendations were proposed, the most important of which is working on training of trainers and those who are in charge of training process on the topics and competencies revealed by the current study through meeting training needs and encourage training managers or planners to recruit change leadership in training and utilizing of modern training methods in training and activation of strategic planning for training.

Study [4] aimed to identify the role of tacit knowledge in development of human resources in multinational companies under the concept of globalization by identifying the extent of awareness of employees in a multinational company in Jordan to use the tacit knowledge and its characteristics. The impact of that on human resources development under the concept of globalized management, obtaining the capacity to set a proposal for development of human resources by using tacit knowledge approach in these companies according to the concept of globalized management.

The study question is an attempt to answer the following question: is tacit knowledge has a role and impact in human resource development under the

concept of globalized management in multinational companies? The researcher used analytical descriptive method, and the study population is multinational companies in Amman city. The study sample is 15 companies. The study concludes to the following results: the tacit knowledge in terms of usage, application and characteristics influence human resource development and has a role under the globalized management concept in multinational companies, there is a strong positive relationship between the variables of the study, no differences that were due to the variables of the study such as: (Qualifications—years of experience—functional level) between the sample in the extent of using of tacit knowledge, characteristics, development of human resource and the concept of globalized administration.

Study [5] aim to reveal the relationship between tacit knowledge in respect of technical cognitive dimensions from one hand, and the quality level of Palestinian telecommunication firm products with tangible physical and intangible ingredients, which are the performance and conformance quality (compliance), durability, validity, complementary characteristics, reliability, aesthetic qualities and perceived quality. Therefore knowing whether there was a mutual influence between tacit knowledge and the level of product quality, and the level of this impact, as well as how it can be strengthened in case of a positive or improve it if it was weak. The study concludes to the result of this spiritual relationship from a statistical view, and recommended more in-depth studies in this field.

3 PROBLEM FORMULATION

The question of this study came from here, to learn about the reality of these tacit knowledge sources in the field of education that offered through educational training centers, especially Educational Training Center in the Department of Education in Bisha province. The study will try to answer the main question, which states, “what is the reality tacit knowledge sources in Educational Training Center in the Department of Education in Bisha province?” and the three sub-questions of the main question that state; “what is the role of tacit knowledge sources in the Educational Training Center in the Department of Education in Bisha province?” as well as the second question which states; “what is the role of tacit knowledge sources in professional development in Educational Training Center in the Department of Education in Bisha province?”, and finally the third question, which states, “what is the role of knowledge technologies in the development of tacit knowledge to develop professional performance in Educational Training Center in the Department of Education in Bisha province?”.

Table 1. Number of respondents and their details.

Training and scholarship director	Head of the training center	Educational training supervisors	Total
1	1	3	5

4 RESEARCH METHODOLOGY AND DATA SAMPLES OF THE STUDY

The researcher used in this research descriptive case study in order to answer the question, and the research tools are interview that the researcher used to answer the first sub-question, and questionnaire are used by the researcher to answer second and third sub-questions, samples is composed of five individuals, including training and scholarship director, the head of the training center and three educational training supervisors as shown in Table 1.

5 NUMERICAL RESULTS

To answer the first sub-question of the research, the researcher using interview technique as follows: the first sub-question stipulates; what tacit knowledge sources in the Educational Training Center in the Department of Education of Bisha province? In order to answer it, the researcher visited the Educational Training Center in Bisha province, applying interview technique where he held interviews with:

- Educational Training Department and scholarship director, Mr. Abdullah Mohammed Al Korkman.
- Director of Educational Training Center, Mr. Nasser Aballah al-Jahmi.

The researcher discusses in both interviews the tacit knowledge sources that benefit and contribute to training centers in training field by feeding knowledge and researcher concludes the tacit knowledge sources that explained in the Fig. 1.

- Descriptive statistics of study sample data

Distribution of respondents according to dealing with computer skills

Analysis: the above table indicates that proportion of excellent dealing with computer skills represent 80.0% of the total study sample who are the majority, while proportion of good dealing with skills computer represent 20.0% of the total study sample.

Analysis:

- Is clear from the above table that the entire study sample represented a rate of 100.0%, they have the skills of dealing with the search engines, and with the databases, and with e-mails.

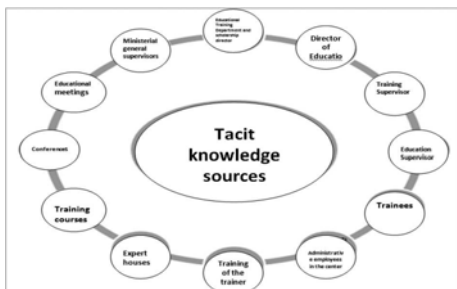


Figure 1. Tacit knowledge sources.

Table 2. Distribution of respondents according to dealing with computer skills.

Dealing with computer skills	Frequency	Percentage %
Non	–	–
Acceptable	–	–
Good	1	20.0
Excellent	4	80.0
Total	5	100.0

Analysis

– Is clear from the above table that the entire study respondents represented a rate of 100.0% who use social networks (Twitter and Facebook, and WhatsApp, and YouTube), and 60% of the total study respondents use blogs and other social networks such as (Instagram, Netlog and ...).

Reliability Coefficient of the study was conducted by using Cronbach's alpha for each axis of the study axes and study as a whole, as shown in Table 5 below:

Considering the results of the test, we find that:

Reliability coefficient values of Cronbach's alpha for axes are very high and close to one, and the reliability coefficient value of alpha Cronbach of the study as a whole is very large and close to one, that means that the questionnaire has a very high Reliability in the field of study application.

Self-validity of questionnaire:

Self-validity was calculated the by using the following equation:

$$\text{Self-validity} = \sqrt{\text{Reliability}}$$

$$\text{Self-validity} = \sqrt{0.9642} = 0.98$$

It is a very high coefficient of validity and statistically significant refers to the possibility of trust in the results that might emerge from the questionnaire.

Table 3. Distribution of respondents according to the deal with the online skills.

Type of dealing skills	Options	Percentage of respondents %
Dealing with search engines	5	100.0
Dealing with databases	5	100.0
Dealing with e-mails	5	100.0
Total	15	300.0

Table 4. Distribution of respondents according to using of social networks.

Type of social networks	Options	Percentage of respondents %
Twitter	5	100.0
Facebook	5	100.0
WhatsApp	5	100.0
You Tube	5	100.0
Blogs	3	60.0
Others (such as Instagram, Netlog and...)	3	60.0
Total	26	520.0

Table 5. Measuring of reliability and validity of the tool.

Axis	Number of paragraphs	Cronbach's alpha coefficient
First axis (the role of tacit knowledge in forming knowledge)	12	0.8622
Second axis (the role of tacit knowledge in sharing knowledge)	12	0.8790
Third axis (the role of tacit knowledge in applying knowledge)	12	0.9358
Fourth axis (the role of tacit knowledge in developing tacit knowledge)	22	0.9433
The entire study coefficient	58	0.9642

Answer to second and third sub-questions

Answer to the second sub-question, which states "what is the role of tacit knowledge sources in professional development in Educational Training Center in the Department of Education in Bisha province?", to the answer to this statistical technique has been used which is represented by arithmetic average and (t-test), One-Sample T Test. At the level of significance: $\alpha = 0.05$.

First, axis of role of tacit knowledge in forming knowledge for professional performance development: The results were as illustrated in Table 6 below:

The results show the following:

The significance value (P. value) = 0.016 which is less than the value of ($\alpha = 0.05$), and this means that tacit knowledge sources have a role in forming knowledge for professional performance development in the Educational Training Center in Bisha province, as well as the results show, that the arithmetic average is equal to (2.80), this means that the tacit knowledge sources have a role by 70.00% in forming knowledge for professional performance development in the Educational Training Center in the Department of Education in Bisha province.

Second, role axis of tacit knowledge sources in sharing knowledge for the professional performance development. The results as illustrated in Table 7:

The results show the following:

The significance value (P. value) = 0.011 which is less than the value of ($\alpha = 0.05$), and this means that tacit knowledge sources have a role in sharing knowledge for professional performance development in the Educational Training Center in Bisha province, as well as the results show, that the arithmetic average is equal to (2.90), this means that the tacit knowledge sources have a role by 72.50% in sharing knowledge for professional performance development in the Educational Training Center in the Department of Education in Bisha province.

Third, role axis of tacit knowledge sources in applying knowledge for the professional performance development. The results as illustrated in Table 8:

The results show the following:

The significance value (P. value) = 0.033 which is less than the value of ($\alpha = 0.05$), and this means that tacit knowledge sources have a role in applying knowledge for professional performance development in the Educational Training Center in Bisha province, as well as the results show, that the arithmetic average is equal to (3.20), this means that the tacit knowledge sources have a role by 80.00% in applying knowledge for professional performance development in the Educational Training Center in the Department of Education in Bisha province.

Forth, role axis of tacit knowledge sources as general professional development in the Educational Training Center in the Department of Education in Bisha province.

The results as illustrated in Table 9.

The significance value (P. value) = 0.016 which is less than the value of ($\alpha = 0.05$), and this means that tacit knowledge sources have a role in professional performance development in the Educational Training Center in Bisha province, as well as the results show, that the arithmetic average is equal to (2.80), this means that the tacit knowledge sources have a role by 70.00% in professional performance development in the Educational Training Center in the Department of Education in Bisha province.

Table 6. Role of tacit knowledge sources in forming knowledge, t-test results.

Axis	Average	Freedom degrees	Calculated (T) value	P. value (sig)
Role of tacit knowledge in forming knowledge for professional performance development	2.80	4	4.00	0.016

Table 7. Role of tacit knowledge sources in knowledge sharing, t-test results.

Axis	Average	Freedom degrees	Calculated (T) value	P. value (sig)
Role of tacit knowledge in sharing knowledge for professional performance development	2.90	4	4.20	0.011

Table 8. Role of tacit knowledge sources in knowledge applying, t-test results.

Axis	Average	Freedom degrees	Calculated (T) value	P. value (sig)
Role of tacit knowledge in applying knowledge for professional performance development	3.20	4	3.207	0.033

Table 9. Role of tacit knowledge sources in professional performance development, t-test results.

Axis	Average	Freedom degrees	Calculated (T) value	P. value (sig)
Role of tacit knowledge sources in professional performance development	2.80	4	4.00	0.016

Table 10. Role of tacit knowledge technologies in professional performance development, t-test results.

Axis	Average	Freedom degrees	Calculated (T) value	P. value (sig)
Role of tacit knowledge technologies in professional performance development	2.80	4	4.00	0.016

Third sub-question, which states “what is the role of knowledge technologies in the development of tacit knowledge to develop professional performance in Educational Training Center in the Department of Education in Bisha province?”, to the answer to this statistical technique has been used which is represented by arithmetic average and (t-test), One-Sample T Test. At the level of significance: $\alpha = 0.05$. The results as illustrated in Table 10.

The significance value (*P. value*) = 0.016 which is less than the value of ($\alpha = 0.05$), and this means that knowledge technologies have a role in professional performance development in the Educational Training Center in Bisha province, as well as the results show, that the arithmetic average is equal to (2.80), this means that the tacit knowledge sources have a role by 70.00% in professional performance development in the Educational Training Center in the Department of Education in Bisha province.

Determining the role of each technique of knowledge technologies in developing of tacit knowledge for professional performance development in the Educational Training Center in the Department of Education in Bisha province based on respondents views.

Role of all technologies in developing of tacit knowledge for professional performance development.

From results we find that search engines role has a role in formation of knowledge with 90.00% and sharing of knowledge with 75.00%, and databases has a role in formation of knowledge and sharing of knowledge with 70.00%, as well as emails contribute to formation of knowledge and sharing of knowledge with 75.00%. While in application of knowledge contribute with 70.00%. Twitter also contributes to formation of knowledge and in

application of knowledge with 70.00%, as well as in sharing of knowledge with 75.00%. Facebook contributes to formation of knowledge and sharing of knowledge with 65.00%; while in application of knowledge contribute with 70.00%. whatsApp contributes in formation of knowledge with 70.00%, And contributes to sharing of knowledge with 75.00%; while in application of knowledge contribute with 60.00%. YouTube contributes to formation of knowledge and sharing of knowledge with 75.00%, and in application of knowledge contributes with 70.00%. Blogging also contributes to formation of knowledge and sharing of knowledge with 75.00%, while in application of knowledge contributes with 65.00%. Finally, and in generally, we find that knowledge technologies have a role in developing of tacit knowledge for professional performance development with 70.00%.

6 CONCLUSION AND RECOMMENDATIONS

As for the question “what is the role of tacit knowledge sources in professional development in the Educational Training Center in Department of Education in Bisha province?” by using a t-test (T) One-Sample the researcher finds that tacit knowledge sources have a role in knowledge formation for professional performance development in Educational Training Center in Bisha province, with arithmetic mean of (2.80), therefore tacit knowledge sources have a role with 70.00% in knowledge formation for professional performance development in Educational Training Center in Bisha province. Tacit knowledge sources have a role in sharing knowledge to develop professional performance in the Center Educational Training in Department of Education in Bisha province, with

arithmetic mean of (2.90); therefore tacit knowledge sources have a role with 72.50% in knowledge sharing for professional performance development in Educational Training Center in Bisha province. Tacit knowledge sources have a role in application of knowledge to develop professional performance in the Center Educational Training in Department of Education in Bisha province. Tacit knowledge sources have a role in applying knowledge to develop professional performance in the Center Educational Training in Department of Education in Bisha province, with arithmetic mean of (3.20), and tacit knowledge sources have a role with 80.00% in knowledge application for professional performance development in Educational Training Center in Bisha province. Finally, as general we find that the tacit knowledge sources have a role in developing of professional performance in the Center Educational Training in Department of Education in Bisha province, with arithmetic mean of (2.80), and tacit knowledge sources have a role with 70.00% in professional performance development in Educational Training Center in Bisha province.

As for the question “what is the role of knowledge technologies in the development of tacit knowledge to develop professional performance in Educational Training Center in the Department of Education in Bisha province?”, by using a t-test (T) One-Sample the researcher finds that tacit knowledge sources have a role in developing tacit knowledge for professional performance development in Educational Training Center in Bisha province, with arithmetic mean of (2.80), therefore knowledge technologies have a role with 70.00% in developing tacit knowledge for professional performance development in Educational Training Center in Bisha province.

As well as databases have a role in formation of knowledge with 70.00%, and to knowledge sharing with 70.00%. Emails contribute to formation of knowledge with 75.00%, and to the sharing of knowledge with 75.00%, and to application of knowledge with 70.00%. Twitter contributes to formation of knowledge with 70.00%, and in sharing of knowledge with 75.00%, and in application of knowledge with 70.00%. Facebook contributes to formation of knowledge with 65.00%, and to sharing of knowledge with 65.00%, and to application of knowledge with 70.00%. WhatsApp contributes to formation of knowledge with 70.00%, and to sharing of knowledge with 75.00%, and to application of knowledge with 60.00%. YouTube contributes to formation of knowledge with 75.00%,

and to sharing of knowledge with 75.00%, and to application of knowledge with 70.00%. Finally, Blogs contribute to formation of knowledge with 75.00%, and to sharing of knowledge with 75.00%, and to application of knowledge with 65.00%.

This study suggests working on creating a culture of sharing and sharing of knowledge in the field of educational training in general and training in particular centers, in addition to working on the documentation of seminars, lectures, courses and where, organizing and storing what has been documented and made available and shared with interested each area communities, and finally go about trading sharing knowledge using modern technologies, especially social media.

This study recommends conducting researches specializes in studying tacit knowledge sources in field of educational supervision, and conduct researches specializes in studying tacit knowledge sources in field of students counseling, as well as conduct researches specializes in studying tacit knowledge sources in field of educational management. Finally, it recommends conducting research specializes in studying tacit knowledge sources in field of school administration.

REFERENCES

- [1] Asiri, Mahmoud (2011), “The effectiveness of the educational training centers in development of management performance for managers of primary schools in Jeddah, from their point of view.” Unpublished research project, the Higher Educational Studies Program, King Abdul Aziz University.
- [2] Al-Yawar, Afaf Ben Salah Hamdi. (2007). Educational Training in light of contemporary transformations. Cairo, Dar Al Fik Elarabi.
- [3] Abbasi, Mahmoud Salah Suleiman. (2013). “Training needs in training centers in light of contemporary trends in Mecca”. Master thesis, Um al-Qura, Mecca.
- [4] Alsalih, Asmaa Rashad Nayef. (2012). “Tacit knowledge and its role in development of human resources in light of the concept of globalized management: Applied Study on multinational companies.” International Scientific Conference on globalization of management in the era of knowledge 15–17/December 2012, Tripoli, Lebanon.
- [5] Khasib, Khalid, Abu Fada and Marwan. (2012). “Tacit knowledge and its relation with product quality: An Empirical Study on Palestinian telecommunications companies.” Economic Conference of the Al Quds Open University about toward strengthening competitive of Palestinian products, 16–17/10/2012. Ramallah, Palestine.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Digital institutional repositories in Saudi universities between facts and future: A comparative study between Australian and Saudi digital institutional repositories

Khalid Bawazeer

Department of Information Science, Faculty of Arts and Humanities, King Abdul-Aziz University, Jeddah, Saudi Arabia

ABSTRACT: The digital institutional repositories are new positive trend in the field of scientific communication between researchers all over the world, and due to the need to access a wider scientific data; repositories considered an important means and promising to gain access to that data, increased scientific publishing, and enhance communication among researchers.

The importance of the study is to find the reality of digital repositories in Saudi universities and what is the foreseeable future to it. The study aimed to identify the reality of digital institutional repositories in Saudi universities through identifying the strengths, weaknesses, the opportunities and threats by using SWOT analysis approach then compare it with Australian digital institutional repositories. By using the descriptive method our finding clarifies that Saudi Digital institutional repositories are very few. The existing infrastructures not utilized, Staff not trained, Authors are not motivated. On the other hand, the government support is available. There are some universities communicated with international institutional repositories. The copyright polices not effective. Materials are limited.

Keywords: digital, institutional, repositories, infrastructure, software, staff, SWOT

1 INTRODUCTION

The enormous development of technology in this century and intellectual growth of production made it necessary to have digital institutional repositories to accommodate the increasing intellectual production. European countries started ahead of the Arab countries in this field.

Digital Institutional repositories are gaining great importance in the universities and research centers because they provide the potential to save digital contents and manage them, broadcast, and allow the exchange of information and expertise on a local, regional and global level, contributing to the development of courses operations. They will form an integral part in regional blocs between libraries and universities. (Schöpfel, 2009).

The digital repositories is the latest digital information institutions on the Internet, these warehouses featured within the framework of free access to information initiatives, and the most famous digital repositories of institutional follow Universities, scientific or research organization. The availability of intellectual production workers and scientific institution in the digital online free. (Majmaah Univ., 2016)

Throughout this study, the researcher will try to find the status of Saudi digital institutional repositories and what could be the future for them. To reach the main findings, previous Australian and Saudi digital repositories in universities studies will be used and will be analyzed by using a descriptive method and SWOT analysis tool, then the researcher will state the conclusions and recommendations for the future.

2 PREVIOUS STUDIES

The studies in digital institutional repositories are varies on various aspects and different locations. For the purpose of this study, the researcher selected specific previous studies to compare between Saudi and Australian digital institutional repositories in universities as follows:

- (Alyateem, A. & Bn Hameed, Nawaf. 2015) “Digital repositories in the Arab universities: A comparative analytical study”. This study aims to analyze and compare the knowledgeable repositories in some Arab universities. The importance of this study stemmed from showing us the advantages and the disadvantages of the

digital repositories in the Arab universities and it sheds light on its excellent sides for supporting it and on the defects of its treatment then avoided it in the future. By applying SWOT analysis on this study the Strengths: the main languages such as English, Arabic are used in institutional repositories. Weaknesses: the institutional repositories policies are not clear in most of repositories in the Arab universities, there are a few number of institutional repositories. Threats are the cost of the programs that are increasing and technologies are growing rapidly. Opportunities: to utilize the avail universities infrastructures. Therefore, the above analysis concluded that the Arab government supports the institutional repositories is very weak and the available materials are limited need to be increased in order to improve the interaction between institutional repositories.

- (Faraj, 2012). The role of Digital Institutional repositories in supporting and enriching the Arabic content on the Internet. The problem of this article is to identify the reality of the digital institutional repositories and identify the shortcomings of these institutions. In addition, the article analyzes the current situation of institutional repositories in the Arab countries and provide an indication of shortcomings and weaknesses in those locations then offer some suggestions. By applying SWOT analysis on this study, the Strengths are: the existence of policies of copyrights, multi new technologies help in sharing resources among the growing number of digital repositories. Weaknesses are: the small number of digital repositories in the Arab countries, shortage of materials available and the weakness of the content shortage of motivation. Opportunities are: utilization of the government support to universities. Threats are: the digital scientific revolution, the digital gap is widening between the Arab countries and developed countries. Therefore, the above analysis concluded that the Arab countries should increase the number of digital repositories, activate a stimulating policy for deposit, and support them to keep up with scientific and technological revolution.
- (Kemman, M.A. & Kimsley, D., 2009) "The State of the nation: a snapshot of Australian institutional repositories", the study aims to provide a 'snapshot' of the state of Australian institutional repositories as at September 2008. In doing so, it builds on similar research canvassing institutional repositories in particular countries such as the United States, Canada, Britain and France. By applying SWOT analysis on this study, we find that the Strengths are: the Australian government supports the development of institutional repositories, availability of the

infrastructure in institutional repositories in universities. On the other side, the main weaknesses are: the operational costs is high, limited number of programed used. Threats are the funding is not secure for their repository staff, repositories programs are growing rapidly. Opportunities are to utilize the avail programs, to be a resource to other digital repositories. Therefore, the above analysis concluded that the Australian government supports the institutional repositories and the used programs are limited and need to be increased.

3 THE STUDY PROBLEM

Despite the importance of the digital institutional repositories in educational institutions, but the status of the digital institutional repositories in Saudi universities is not clear to the researcher so the formulation of the problem of this study will be: "What is the status of digital institutional repositories in Saudi universities and what could be the future?".

4 METHODOLOGY OF SOLUTION

For the purposes of this study the researcher will used the descriptive and analytical approach, in addition to the review of intellectual production in Arabic and foreign languages to learn about trends in research in the field of digital institutional repositories.

The study hypothesis are: is there relation between the available material in digital institutional repositories and copyright policies.

The study boundaries are Saudi and Australian universities in 2015.

The study tools used is SWOT Analysis. The method used in the analysis in various areas, a tool used in strategic planning to assess the strengths and weaknesses, opportunities and threats points. Strengths and Weaknesses of the internal characteristics of the project. Opportunities and Threats they have on the external conditions for the project.

The digital institutional repositories in Saudi universities are located in King Fahad University, King Saud University Naif University, king Abdul-Aziz University and Majmaah University. Through the next study, the researcher will use SWOT analysis in detail.

(Aldhuwaihy, Fahad, 2014) "Institutional digital repositories in Saudi Universities: Towards a National Project Vision to support the initiatives of its establishment and its management". The study deals with the institutional repositories in

Saudi universities, in terms of lack of digital institutional repositories and slow in development in Saudi universities. The study aim primarily to come up with a vision for a national project to support the creation, management and development of digital institutional repositories in Saudi universities. By applying SWOT analysis on this study, the Strengths are: the higher management supported projects, university budget are available. Weaknesses are: lack of institutional repositories policies, a few number of institutional repositories and material. Threats are: lack of trained staff, technologies are growing rapidly. Opportunities are: to utilize the available universities infrastructures. Therefore, the above analysis concluded that the Saudi government supported the institutional repositories and low motivation to the authors to provide materials, use limited software and limited access to the material.

5 MAIN FINDINGS

The study used the descriptive method and by comparing between Digital institutional repositories in Saudi universities with Australian universities, the main findings are: digital institutional repositories in Saudi universities are very few. The existing universities infrastructures are not utilized. Staff are not trained well. Authors are not motivated to interact with repositories. Government support is available. Some Saudi universities communicated with international digital institutional repositories. The copyright polices are not effective. Materials are limited.

6 CONCLUSION AND RECOMMENDATIONS

The researcher started this study by presenting the problem of the study in a question format, which was “What is the status of digital institutional repositories in Saudi universities and what could

be the future?”. Then SWOT analysis were used to different previous studies then the conclusions of this study are: the digital institutional repositories in Saudi universities are very few and need to be supported by the higher management in the universities.

Finally the recommendations of this study are: start establishing digital institutional repositories, complete the universities infrastructure, train competitive staff, communicate with international institutional repositories to utilize their material, motivate authors and users to interact with digital institutional repositories, establish apply affective polices, easy processes to the users to interact with it.

REFERENCES

- Aldhuwaih, Fahad. (2014). Institutional digital repositories in Saudi Universities: Towards a National Project Vision to support the initiatives of its establishment and its management, available at: https://graduatestudies.kau.edu.sa/Show_Res.aspx?Site_ID=306&LNG=AR&RN=65832 (10/02/2016).
- Alyateem, A. & Bn Hameed, Nawaf. (2015). Digital repositories in the Arab universities: A comparative analytical study, International Conference on Communication, Management and Information Technology (ICCMIT 2015), Available online at: www.sciencedirect.com (02/02/2016).
- Faraj, hanan ahmed. (2012). The role of Digital Institutional repositories in supporting and enriching the Arabic content on the Internet. King Fahd National Library Journal, vol 18, p 2. Available at: www.kfnl.org.sa/Ar/mediacenter/EMagazine/DocLib/.../93-132.pdf (01/02/2016).
- International digital repositories, (2016). Majmaah university, available at: <https://www.mu.edu.sa>. (03/02/2016).
- Kemman, M.A. & Kingsley, D. (2009). The State of the nation: a snapshot of Australian institutional repositories. Available at: <http://firstmonday.org/ojs/index.php/fm/article/view/2282/2092> (01/02/2016).
- Schöpfel, J. (2009). Grey literature in French digital repositories: A Survey. *Conference Papers: International Conference on Grey Literature*, pp. 39–53.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

The role of digital repositories in supporting the main functions of universities: A survey of Arab universities

Nasser Juwaber AlKhudairi

Department of Information Science, Faculty of Arts and Humanities, King Abdul-Aziz University, Jeddah, Saudi Arabia

ABSTRACT: This article aimed to highlight the role of digital repositories in supporting the three main functions of universities: teaching, scientific research and community service. This work focuses on studying and highlighting that role in the Arabic universities. The importance of this goal emanates from its association with three key functions entrusted with the universities, as well as its association with very important facilities represented in Arab universities. To achieve this goal, the researcher selected four previous studies and used analytical method of SWOT analysis that is used to identify the strengths and weaknesses as well as opportunities and threats. The study found that digital repositories contribute to change the culture of teaching through facilitating constant evaluation by lecturers for the performance of their students, as well as gives them opportunity to focus on improving the educational process. The digital repositories also support cooperation and partnership between academic departments to get the scientific output of research that allows partnership to develop and ensure its quality. In addition, it contributes to the distribution of information to the community and delivers that information at fast pace and less cost.

Keywords: Digital Repositories, Institutional Digital Repositories, Functions of Universities, Teaching, Scientific Research, Community Service and SWOT

1 INTRODUCTION

The digital repositories are considered one of the most important digital information institutions on the World Wide Web (Internet). These repositories emerged within the framework of free access to information. One of the well known kinds of these repositories is “institutional digital repositories” which usually affiliates with a university and scientific or research organization to preserves the intellectual property of those universities or organization [1].

The digital repositories are a system for storing content and intellectual assets and preserve it digitally in order to search for and retrieve it when needed. Digital repositories allow the import and export of these assets, their identification, storage and retrieval. The digital repositories are a type of system to manage the content that combines the intellectual assets of the institution and allows its use to support many of the activities within the institutions to which it is affiliated [2].

The institutional digital repositories are a database available on internet. They include digital intellectual production that are deposited by researchers or issued by the institution that

is usually affiliated with universities, institutes and research centers without impediments and restrictions. Also, it may include other types of institutions, such as government departments, agencies, federations’ associations and commercial entities that wish to preserve and distribute their products for free [3].

The most important definitions that are frequent in the literature relevant to the topic which sees that University-Based institutional repository is a range of services offered by university for its members to manage and transmit digital materials created by the university and working associates. More precisely, they are digital archives for the intellectual production of associate members of academics, scholars, researchers and students in the institution. And these digital archives are available for beneficiaries both within the organization and outside. The basic function of the institutional repository is to capture and store researches and all other types of intellectual production issued by the institution to reserve the intellectual rights and publish it on long-term basis [4].

Studies have indicated [1] that the world has witnessed in the recent time remarkable development in building of educational repositories. Canada,

Australia, America and Britain were at the forefront of countries that focused on building these digital repositories and its development. The repository of Kiro (WWW.NETERA.CA) in Canada; Mueller's repository (WWW.MERLOT.ORG) in United States; Edna's in Australia (WWW.EDNA.EDU.AU); and Georm's repository in the United Kingdom (WWW.JORUM.AC.UK), are the most famous repositories that are rich of millions of educational resources.

On the Arab level, the previous studies have confirmed [6] that King Fahd University of Petroleum and Minerals has built a standard digital repository that of its first kind in the Arab world. The digital repository of the library of Alexandria in Egypt also occupied a prominent place among the Arab digital repositories as it has been added to the World Directory of Open Access Repositories which is considered the largest and most important directory for digital repositories.

From here, it is noted that there are obvious weaknesses in the field of digital repositories on the Arab countries level in addition to the weakness of Arabic content on the internet. Therefore, King Abdullah initiative for Arabic content was taken, which aims to enrich and harness the Arabic content to support the development and transition to a knowledge-based society, as well as to ensure that the information is in reach of all segments of the community.

It is well known [5] that the universities have three main functions: teaching, scientific research and community service. Whereas improving the overall performance of the university and to develop its capabilities require improvement in efficiency and effectiveness of the implementation of these key functions, the university always seeks to use all the techniques and tools that help it in achieving that. Therefore, Arab Universities are seeking to catch up with international universities in the creation of institutional digital repositories that contribute to improve their ranking among universities on global level. They are also seeking to take advantage of the digital repositories to save digital content and broadcast it as well as to allow the exchange of information and experiences on both local and global level in addition to contributing to the process of curriculum development. Many studies have emphasized the importance of institutional repositories as they constitute a part of regional blocs between libraries and universities [6].

Given the importance of the role that is played by improving the efficiency and effectiveness of the three main functions in the development and growth of Arab Universities; and the lack of studies that discussed the nature of the role played by digital repositories for the main functions of the universities; and taking into account the recommendations

that were recommended by one of the previous studies [7], the aim of the present study is to clarify the role of digital repositories and the impact of that role on the basic functions of universities and specifically on the Arab universities.

This research includes several sections. It begins by selecting some of the previous studies, which will be analyzed using SWOT approach to identify strengths and weaknesses as well as opportunities and threats that are discussed in these studies. Then the research will identify the problem as well as the methodology and research tools to reach the desired results and the proposed recommendations.

2 LITERATURE REVIEW

Intellectual production is rich of several studies that discussed multiple aspects of digital repositories. And perhaps most of these studies examined the existing repositories, or putting in place mechanisms to build and implement new repositories. By extrapolating this intellectual production, it became clear that this research is a new and unprecedented research. Therefore, it is worthy of study and research. Although there are many studies on digital repositories, but there is a dearth of studies on the role that is played by repositories in the support of key functions of universities. Here are some previous studies relevant with the research topic:

2.1 *Study of Alyateem & Bn Hameed, 2015 [8]*

This study aimed to analyze and compare digital repositories in some Arab universities and clarify the advantages and disadvantages in those repositories. The study used a descriptive analytical method and descriptive comparative method on a sample of digital repositories governed by six Arab universities. These are: Alexandria University, the University of Damascus, Khartoum University, King Saud University, King Fahd University of Petroleum and minerals and Naif University for security Sciences. By applying the analytical method of SWOT on this study, we found that the strengths are as follows:

The digital repository of the University of Alexandria distinguished with huge content of the sources of information and making it available for social share for books, as well as its support for five different languages. King Fahd University and the University of Naif for security Sciences distinguished by advanced search which gives the researcher a high capacity and greater accuracy during the research process.

The weaknesses were the clear absence of policies in the repositories of Khartoum University

and King Fahd University as well as the absence of social networking tools in King Saud University and King Fahd University. On the other hand, the absence of security services is one of the biggest challenges that each of Damascus University, King Saud University, and Naif University for security Sciences are facing.

In terms of the opportunities cited by the study, there is a need to carry out deep studies to build digital repositories in Arab universities and compare it with its counterpart in order to reach to a unified model to build digital repositories in all Arab universities. This study linked directly to the subject of this paper as it studied six digital repositories affiliated with six Arab universities.

2.2 *Study of Jerioa, 2014 [7]*

This study aimed to clarify the concept, characteristics and types of repositories which can be used by a university professor in teaching practices in faculty of education at Princess Nora University. The female researcher used descriptive and analytical method and questionnaires to achieve the goals of the study. The study concluded with a set of results that show, in its entirety, the importance of the use of digital repositories in teaching and educational practices. By applying the analytical method of SWOT analysis, the strengths of this study concentrated in considering digital repositories as one of the important tools in electronic education. The positive impact of the digital repositories represented in the compilation of scattered educational resources and their authentication which will make it easier to consult when needed.

In addition to their composition for valuable cybernetic balance that help employees of academic institutions in their scientific, educational and research works. The weaknesses points highlighted by the study represented in that the teachers do not possess the skills required for the design and use of digital objects repositories in the educational process. The threat cited by the study represented in lack of concept of digital objects repositories to the majority of faculty members as well as deficiencies in their use for teaching practices. In addition to the failure of universities to provide mechanism to maintain and build educational objects repositories in electronic libraries and the lack of sufficient skill to choose the right digital object for educational content. As far as the opportunities concerned, the study recommended to build digital repositories within the electronic libraries in the university as well as to provide a directory of digital repositories to help faculty members on the optimal use of these technologies. This study is closely related to the subject of this paper as it is related to an important educational

facility, Princess Nora University, and studied digital repositories in terms of their use by faculty members.

2.3 *Study of Alarabi, 2011 [6]*

This study aimed to examine and analyze fifty repositories according to the order of Cybermetrics Lab, in addition to distribution of their contents numerically, qualitatively, temporally, and topically, besides identifying methods of search, retrieval, softwares that were used, and the policies pursued in them. The study did that in view to develop a mechanism to create digital repositories and enable Arab universities to take guidance while building digital repositories. The study relied on descriptive analytical method and found some important results that are: the repositories were keen to provide many methods that enable users to retrieve various information, 75% of the repositories used open source software.

Eprints program was one of the most frequently used programs by 45.8%. The repositories covered in this study were also keen to make a policy for them by 79.2%. By applying the analytical method of SWOT analysis, the strengths of this study concentrated on the grounds that the digital repositories contribute to change the culture of teaching, scientific research and the distribution of information to the community and deliver them at fast speed and less expense. Moreover, the digital repositories represent universities' activities globally through digital availability of the information they provide. The weaknesses highlighted by the study concentrated in the absence of the Arab countries on the list of the best fifty repositories in the world, except Saudi Arabia, which contributed with the depository of King Fahd University for Petroleum and Minerals. The opportunities highlighted by the study are that to allow free access to the educational sources contribute to the development of curriculum. Also, the digital repositories highlight university's place and its role in the educational and research process. Lack of interest in Arab universities to create digital repositories is considered one of the threats that lead to the loss of intellectual works of its staff and associates. This study is closely related to the topic of this paper in terms of its examination of fifty digital repositories in the world and trying to prepare a mechanism to establish digital repositories in Arab Universities.

2.4 *Study of Monge, Ovelar, & Azpeitia, 2008 [9]*

This study mentioned that repositories of educational objects constitute a comprehensive strategy to support the use of information technology in

educational contexts. From this perspective, the participation of users in repository services is a desirable goal. The study suggested six strategies for Web 2.0 that are used in promoting social dynamic and participation in the repository. By applying method of SWOT analysis on this study, the strengths found in this study concentrated in that participation strategy through the web 2.0 increases the participation of users in educational objects repositories. Social software leads to innovation in educational resources and practices.

The concept of the repository 2.0 is designed to get the maximum benefit from the content generated by the user. The weaknesses highlighted in the study relates to the development of education classifications (educational taxonomies) for use in the recommendation to users while interring of social labels (social tagging input). It is necessary to study the factors of success in various authoring tools for the implementation of most successful authoring aspects for repositories and rapid content creating tools. The opportunities lie in enabling non-identified users to read the repository content. The identified users are such as teachers, who contribute to the repository through developing new educational content and reviewing the content submitted by other users. As well as the opportunities lie in encouraging administrators who publish high-quality content with economic gifts and incentives. This study directly linked to the topic of this paper through its study of educational repositories and strategies of Web 2.0 which allows social participation.

3 PROBLEM STATEMENT

The problem of the study lies in the lack of clarity of the role of digital repositories in support of the three main functions of the universities (teaching, scientific research and community service) in Arab universities. This problem clearly emerges in the lack of interest of Arab universities, until recently, in establishing digital repositories to manage their contents and academic assets and make it available to faculty members and students. Therefore, this paper attempts to answer the following question: "What is the role of digital repositories in support the three main functions of the universities in Arab universities." To answer this question, we must answer the following sub-questions:

- What is the role of digital repositories to support teaching job in Arab universities?
- What is the role of digital repositories in supporting scientific research process in Arab universities?
- What is the role of digital repositories in support the community service activity in Arab universities?

4 THE PROPOSED METHOD

This study used a survey method with a comprehensive collection of information that relate to the topic of the current study and to the role played by digital repositories in support of the main functions of the universities in Arab universities. The study used a quadruple analysis tool called SWOT. It is a general strategic analysis tool used in many areas of analysis and aims to highlight the strengths and weaknesses. It also helps to focus on the opportunities and threats.

The SWOT analysis emerged as a result of the research that was carried out in the Institute of Stanford from 1960 to 1970 by Albert Humphrey and other associates of the Institute. The purpose of conducting that research was to find out the reasons for failure of corporate planning and the economic problems that emerged out from this failure, in addition to know ways to avoid these problems. This method is used in the analysis in various areas. It is a tool used in strategic planning to assess the strengths and weaknesses, opportunities and threats points.

The strengths and weaknesses points represent the internal characteristics of the project. While the opportunities and threats represent the external conditions for the project. This study sample is concentrated in that the researcher chooses a specific practical sample represented in the three main functions of the University. He carried out a study about the role of digital repositories in the efficiency and effectiveness of the implementation of these functions in Arab universities. The limits of the objective study lie in the three main functions of universities. While the spatiality limits of the study lie in Arab universities. And the temporal limits cover the second semester of the year 2015 and the time period to cover the studies and scientific research was between the years 2008 to 2015.

5 RESULTS AND DISCUSSION

After reviewing the previous studies and applying SWOT analysis method, the researcher found a number of important outcomes that determine aspects and trends that supported digital repositories to raise the efficiency and effectiveness of the implementation of the three main functions of universities that are: teaching, research, and community service. The following points represent the findings of the study:

First: Aspects supported by digital repositories for teaching in universities

1. Availability of its contents for all beneficiaries of the students and faculty members without any obstacles or restrictions.

2. Contribute to the process of curriculum development.
3. Contribute to obtain information necessary for the development of curriculum.
4. Contribute to changing the culture of teaching through facilitating evaluation by lecturers for the performance of their students and giving them opportunity to focus on improving the educational process.
5. It contains several types of materials that can be in digital form from the beginning, or converted to a digital form. This will enrich and harness teaching job in universities.
6. It is distinguished with excellence and continuity in the content that allows teachers and students to get directly aware of the general intellectual output of the University.
7. It supports individual responsibility for what is deposited in the content and this enhances the skills and quality of work for students and professors.
8. It save the time of beneficiary in getting special educational sources for curriculum.

Second: Aspects supported by digital repositories for scientific research in universities

1. Possibility of saving special intellectual content of the university as well as its management and broadcast.
2. It supports cooperation and partnership between academic departments to get the scientific output of research and etc. that allows partnership to develop digital repositories and ensure their quality.
3. It takes care of intellectual property rights of the electronic content and contributes to the reduction of costs of publishing and printing as well as assist in achieving concept of paperless society.

Third: Aspects supported by digital repositories for community service at universities

1. Possibility of long-term saving of research and intellectual assets and take advantage of it in support of development plans in the community.
2. It allows free access to the sources for the beneficiaries from outside the university.
3. It distributes information and delivers them to the community quickly at less cost.

6 CONCLUSION AND RECOMMENDATION

The study sought to highlight the role of digital repositories in support of the three main functions of universities, namely, teaching, scientific research, and community service. The study found significant results devoted to the importance

of digital repositories in support of the main functions of universities. From here, it is clear that Arab universities should speed up to keep pace with western universities to establish institutional digital repositories that help to accomplish basic functions assigned to universities. In light of the above results, we can make the following recommendations:-

1. The establishment and development of educational digital repositories in Arab universities should be expanded in the light of the quality standards.
2. Universities are sought to strive towards establishing cooperative framework for establishment of educational digital repositories.
3. We should harness the skill of faculty members to create digital content for repositories and use it in various educational aspects.

REFERENCES

- [1] Omar, Iman Fawzi, (2011). Origins and Evolution of Open Digital Repositories. *Cybrarians Journal*. Retrieval date 4.2.2016, from http://journal.cybrarians.info/index.php?option=com_content&view=article&id=607:2011-12-02-01-38-43&catid=252:2011-11-28-21-19-07.
- [2] Hayes H. (2055). Digital Repositories: Helping universities and colleges. Retrieval date 5.2.2016, from [http://www.jisc.ac.uk/uploaded_documents/JISC-BP-Repository\(HE\)-v1-final.pdf](http://www.jisc.ac.uk/uploaded_documents/JISC-BP-Repository(HE)-v1-final.pdf)
- [3] Faraj, Hanan Ahmed (2012). Institutional digital repositories and its role in the support of Arabic content and its enrichment on the Internet. *King Fahd National Library magazine* 94–132. Retrieval date 5.2.2016, from <http://www.kfnl.org.sa/Ar/mediacenter/EMagazine/DocLib/%D8%A7%D9%84%D8%AB%D8%A7%D9%85%D9%86%20%D8%B9%D8%B4%D8%B1/93-132.pdf>
- [4] Al-Bassam, Areej Abdullah, Al-Yami, and Huda Yahya (2013). Digital Repositories (LOR) to ensure the quality of e-learning content. *E-learning and distance education*, 1–25. Retrieval date 13. 2.2016 from http://eli.elc.edu.sa/2013/sites/default/files/abstract/rp97_0.pdf
- [5] Ministry of Higher Education (2013), The third function of universities, General Directorate of Planning and Statistics, 1–32. Retrieval date 6.2.2016, from <https://www.moe.gov.sa/ar/Ministry/Deputy-Ministry-for-Planning-and-Information-affairs/The-General-Administration-of-Planning/Documents/2222.pdf>
- [6] Al-Arabi, Ahmed Obada (2011). Digital repositories for academic institutions and their role in the educational process and research and preparing a mechanism to create a digital repository of Arab universities. *King Fahd National Library Journal*, 149 to 194.
- [7] Al-Jerioa, Siham Salman (2014). Use of educational digital repositories in teaching practices for faculty members in the department of education, Princess

Nora Bint Abdul Rahman University. Specialized International Journal of Educational, 114 –133.

- [8] Aal Yateem, A.A., & Bn Hameed, N.B. (2015). Digital repositories in the Arab universities: A comparative analytical study. International Conference on Communication, Management and Information Technology (ICCMIT 2015), 768–777.
- [9] Monge, S., Ovelar, R., & Azpeitia, I. (2008). Repository 2.0: Social Dynamics to Support Community Building in Learning Object Repositories. Interdisciplinary Journal of E-Learning and Learning Objects, 1–14. Retrieved 2.12.2016, from <http://ijklo.org/Volume4/IJELLOv4p191–204Monge.pdf>

Importance of integrating knowledge management methods and tools to enhance risk management processes. Exploratory study in Saudi Arabia business environment

Salem S. Humaidan

Information Science Department, King Abdul-Aziz University, Jeddah, Saudi Arabia

ABSTRACT: As a matter of fact the momentum behind Knowledge Management (KM) is growing. While it's correct that KM can enhance performance in business, innovation, and knowledge sharing, its major importance may lie elsewhere within Risks Management (RM). Therefore, this paper describes the importance of integration of KM methods and tools to give a further boost to the RM processes. Due to the nature of the subject the researcher used descriptive and analytical approaches. This study has revealed that using KM methods and tools can develop RM. The study findings are valid to all organizations employing KM and RM. Moreover, the findings can be further developed to conceptualize new procedures and methods for KM and RM integration through cooperation between the academic researchers and professional managers. And finally the study concluded that Successful integration of KM and RM can be achieved through a balanced assortment of technology, effective processes of work and adaptive mind set of people.

Keywords: knowledge; Risk; Knowledge management; Risk management; tacit; explicit; knowledge methods; knowledge tools

1 INTRODUCTION

Organizations operate in a web of uncertainty (risks) which ranged from the cluster of natural disasters, unstable business environments, and failures related to human factor, violation of security and financial disorder. The high rates of failure can be reduced if risks are managed in a suitable way. Most of the researches on risk have been concentrated on dealing with the negative rather than positive side. This leads to a narrow vision of business uncertainty [1]. Time and cost elements affect other factors leading to less assurance and ignorance of other issues that could potentially create risks. But this is only one facet of RM which as a divergent domain takes in knowledge from varied sources to apply practices for solving specific problematic areas [2].

With the growing size of organizations and increasing complexity, the need for the effectiveness and efficient KM becomes vital [3]. Hence, KM has received considerable attention in the literature [4]. Initially this paper conducts a comprehensive review of the KM and RM approaches literatures. KM processes were studied and identified. Common methods and tools of KM being used were discussed, which is followed by the examination of the similarities between KM and RM processes.

Previous steps resulted in generating a study framework. This framework was further confirmed using the primary collected data by the means of interviews with expert and practitioners in KM and RM from various industries. At the start there is significant overlap between KM and RM but very few authors have utilized KM philosophy in the RM process. Therefore there is an enormous area between these two aspects of business which is still gray and in its improvement phases.

This research intended to discover the extent of using KM methods and tools to enhance RM, not only from academic point of view, but also through identifying the practitioners and experts perspective. Heisig was the earliest researcher who involved companies and KM practitioners in identification of KM frameworks and actions related with KM [4].

The main objective of this study is to recognize the extent to which merging the methods and tools of Knowledge Management are applicable to the processes and procedures of RM. Also we try to implement the following:

1. Perform a comparison study between KM and RM processes.
2. Identify and examine the tools and techniques of KM which can be used in RM processes.

3. Investigates whether KM methods and tools can be effectively implemented in order to enhance current RM practices.

The researcher used descriptive and analytical approaches to achieve these objectives.

Secondary information collected from different sources and formed the base of the structured interview as a main tool to collect the primary information.

The researcher adopted the list of KM methods and tools which was compiled and approved by the Asian Productivity Organization (APO). 12 each Interview were conducted with expert and practitioners from different companies and different business sectors to get their point of view in usefulness of use KM tools and method to enhance RM processes.

The main results that we achieved it prove that there is a high level of similarity between KM and RM processes. More over the practitioners' answers positively supported the main questions of this paper that KM tools and methods are useful to support and enhance RM processes. Finally, This paper concluded that organizations cannot manage their risks effectively unless they manage their knowledge.

2 RESEARCH PROBLEM

It is a fact that the proper implementation of KM enhances the organizations performance, and it considered as crucial factor in organizations success. Knowledge management has a variety of methods and tools that have been adopted by organizations and practitioners all over the world, those tools and methods contribute to maximize knowledge utilization within the organization. The KM is obviously revealed through its enhancement and support of risk management processes.

Despite of occurrence of RM and KM departments within organizational structures of some organization in Saudi Arabia, but the researcher noticed inadequate use of KM tools and methods to support RM.

Therefore this study will try to highlight the importance of integrating KM methods and tools to enhance and support RM within the organizations in Saudi business environment. Consequently the study objectives will be as follow:

1. Describe the KM and RM processes and perform a comparison desk study,
2. Classify the KM tools and methods that proposed to support RM processes,
3. Discover the contributions of each tool and methods in supporting RM processes, from practitioner's point of view.

3 METHOD AND DATA COLLECTION

To achieve objectives of this paper, researcher will use a descriptive and an analytical approaches, by mean of conducted three phases:

First, perform the disk study by mean of gathering the necessary information and data from literature sources such as books, previous researches, articles and journals. Obtain current theories for both the KM and RM literature. Detailed processes will be analyzed to identify appropriate KM and RM process. Recognize the KM tools and methods that could be used to enhance RM.

Second, primary sources in the form of semi-structured interviews will be also utilized to verify the findings from the literature review. Interviews with representatives from organizations will be conducted to analyze and chat about the findings from the first step, which is the base of forming an interview template. Eleven interviews will be conducted with experienced professionals in KM and RM. The author included in the interviews how do managers deal with risks by using KM practices of brainstorming, lessons learnt and others (which may not be directly apparent to them).

Third, comprehensive discussions and analysis will be followed for the data collected from the literature review and the interviews to explore the validity and similarity of KM and RM process, beside, recognize the KM methods and tools which could boost RM. This step formed a rigorous foundation to base the conclusions on.

Data from the interviews will be combined into a single set instead of separate data sets for the different companies because the focus will be on the effectiveness of using KM methods and tools to boost RM processes instead of comparing them between companies. Therefore, the collective data will clarify the methods and tools of KM which can be used to enhance RM processes within all the companies.

4 KM AND RM DISK STUDY

A Knowledge Management

We always use the word "know", but what does it mean? Most of us think that we have an obvious understanding of the concept, but providing a more exact analysis of it is not easy.

1. *What is knowledge?*

In order to clarify the definition of knowledge, it is vital to understand the related terms; data, information, knowledge and wisdom. First recorded occurrence of Date—Information—Knowledge—Wisdom (DIKW) hierarchy was in 1934 by Thomas Stearns Eliot "Where is the wisdom we have

lost in knowledge? Where is the knowledge we have lost in information?” Those questions come from the poem “The Rock” [5].

Data is raw, with no meaning and significance beyond its existence [6]. Information is the shape of data which carries meaning by way of relational link [7]. Knowledge is defined as the individual’s understanding gained and increased by combining of data, information, experience and the person interpretation [8]. In short, knowledge is collected data and information put into a certain perspective. Figure 1 explains the Intellectual transformational relation through these terms.

2. Knowledge Classifications

A significant contribution has been presented by Ikujiro Nonaka and Hirotaka Takeuchi through presenting their categorization of knowledge. They divided knowledge into two classes: Explicit and Tacit.

- Explicit knowledge is the one which can be captured and kept in the form of documents for record purposes.
- Tacit knowledge by nature resides in the mental power of the individual which can be a person or a group [9].

Moreover, tacit knowledge has two dimensions; the technical dimension is the first one, which includes the kind of informal special skills, referred to as (know how). Cognitive dimension is the second one which is deeply embedded in us and encompasses ideal, value, belief, and mental model. Based on this cognitive dimension of tacit knowledge we perceive the world (Nonaka & Konno, 1998).

3. Definition of Knowledge Management

With the previous explanation of knowledge notion, the concept of KM can be understood as follow. Its definition differs among industries and

organizations with the focus on utilizing knowledge to get competitive advantage regardless of the sector and profitability of organizations.

Knowledge management has been identified through different perspectives. Some authors express it through its functionality. Choucri said; KM is about determining who gets what, when, and how [10]. Some authors perceive it as a process rather than a function. They identify knowledge as an intellectual substance within individual minds and demonstrate in texts and behaviors, rather than a process which restricts its understanding [11]. Chawla & Joshi identify KM as procedures starts from identifying and analyzing of available and necessary information leading to the consequent planning and control of activities to develop and improve knowledge assets [12]. NASA has described KM as the transference of accurate information to the right people at the correct time, while supporting individuals to create and share of knowledge to employ information for evidently performance improvement [13].

At all events, KM is the planning, organizing, directing and controlling knowledge cycle to satisfy business needs in order to gain sustainable competitive advantage.

4. Knowledge Management processes

As per Lottering and Dick (2012) the literatures have recognized more than 160 KM models and frameworks. In an analysis of these models and frameworks, Heisig in 2009 called for the harmonisation of the wide range of diffuse KM terms and concepts in order to standardise and consolidate them. He exposed six KM processes that KM models and frameworks use commonly. They are: create, identify, share, acquire, use and store [14]. Evans, Dalkir and Bidian advanced the Knowledge Management Cycle (KMC) model in their research which contains seven stages: identify, store, share, use, learn, improve, and create [15].

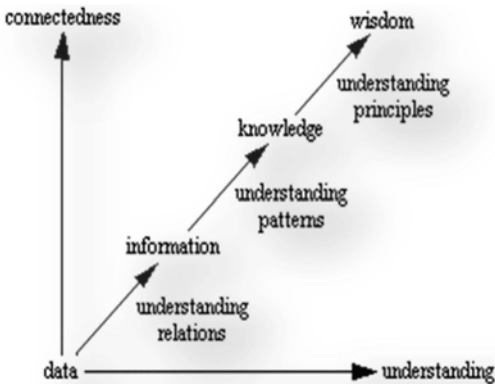


Figure 1. Intellectual transitions [8].



Figure 2. KM processes.

Table 1. KM methods and tools [16].

KM methods and tools	
Ser. Non IT	IT
1 Brainstorming	Document Libraries leading to a Document Management System
2 Learning and Idea Capture	Knowledge Bases (Wikis, etc.)
3 Peer Assist	Blogs
4 Learning Reviews	Social Network Services
5 After Action Review	Voice and Voice-over-Internet Protocol (VOIP)
6 Storytelling	Advanced Search Tools
7 Collaborative Physical Workspace	Building Knowledge Clusters
8 APO Knowledge Management Assessment Tool	Expert Locator
9 Knowledge Café	Collaborative Virtual Workspaces
10 Community of Practice	
11 Taxonomy	

Table 2. Uncertainty facets [18].

Ser.	Type of uncertainty
1	Stochastic Event risk An event it may or maynot happened
2	Aleatoric Variability risk Certain futur event with variable characteristics
3	Epistemic Ambiguity risk Certain future event with ambiguse characteristics
4	Ontological Emergent risks Unknowable unknown

The researcher improved the KMC model by adding organizing phase to the previous model to emphasize the importance of classification and retrieving knowledge.

5. *Knowledge Management Methods and Tools*

The researcher selected the following list of Knowledge Management Methods and Tools which was compiled and approved by the Asian Productivity Organization (APO) KM methods and tools’ expert team in Singapore in August 2009. In order to study the importance of their effect upon risk management processes. These methods and tools were put into practice globally by the most successful organizations, within their KM implementation initiatives [16].

B *Risk Management*

This part provides a review of the Risk Management literature. This is provided by identifying

risk first which followed by explaining the facets of uncertainty. That will follow by classification of risks. The last section will highlight the RM process.

1 *Risk Identification*

Back to 17th century the word risk initiated from Italian word (risciare), which means to dare or to take a choice under unsure conditions. There are many definitions of risk. The definition set out in ISO Guide 73 is that risk is the “effect of uncertainty on objectives”. More detail in risk identification comes from PMI which says risk is “An uncertain event or condition that, if it occurs, has a positive or negative effect on a project objective” [17].

2 *Facets of Uncertainty*

Uncertainty has been divided into four categories; stochastic, aleatoric, epistemic, and ontological [18].

3 *Risk Classifications*

Risks have been classified in many categories such as internal or external risks, Strategic or operational risks. From previous definition which shows two dimensions of risk concept; uncertainty and impact, David Hillson classified risks from the impact side. Risk could be positive (useful) or negative (harmful) [18]. For example; due to labor cost change a project might go over budget as negative risk, the positive risk that the project will be under budget.

4 *Risk Management*

Different organizations and entities define it in different expressions. The Institute of Risk Management perceives it as the process whereby entities methodically concentrate on the risks attaching to their activities with the goal of achieving sustained benefit within each activity and across the portfolio of all activities [19]. Dey described it as “The systematic process of identifying, analyzing and responding to project risk” [20].

No matter what definition is used, the overall idea of RM remains the same which is, All activities, methods and tools used to identify, analyze, plan and implement response to all risks organization may face in order to achieve its goals.

5 *Risk Management processes*

In reference to International Standards Organization, the process of RM contains of five key stages: establish context, identify, analyze, evaluate and respond [21]. However, the researcher revised the ISO model by adding necessary stages and came



Figure 3. RM Processes.

up with a seven steps model for RM process: identify, assess, analyze, plan, implement, monitor and learn as shown in Figure 3.

5 RESULTS AND DISCUSSION

The analysis of the collected secondary data was very vital as it exposed the areas that need to be confirmed through the primary research data in the form of interviews. While doing so, the analysis led to the following results:

1. An eight steps model for KM process was adopted as a result of improving the KMC model by adding an organizing phase to emphasize the importance of the classification and retrieving knowledge. These steps are: identify, produce, organize, store, share, use learn and improve, as shown in Figure 2.
2. The researcher revised the ISO model for RM processes, and came up with a seven steps model. These steps include: identify, assess, analyze, plan, implement, monitor and learn, as shown in Figure 3.
3. By reviewing Figure 2 and 3, it is obvious that every step of RM process is falling within KM process, which means there is a high level of similarity between KM and RM processes. Which answer the first question: Perform a comparison study between KM and RM processes
4. This paper adopted the following list of KM methods and tools which was compiled and approved by the Asian Productivity Organization (APO) KM methods and tools expert team in Singapore in August 2009. As seen in Table 2.1 and Table 1.

With the previous results apparent from the analysis of secondary data, it was essential to build up these results into segments of questions for validation through interviews to verify

the Validity and applicability of KM and RM processes moreover, discussing the usefulness of KM methods and tools. This was arranged in the form of an interview guide (shown below), as a research tool, which was used for every interview.

The interview guide

1. Review RM processes; identify, assess, analyze, plan, implement, monitor and learn.
2. Review KM processes; identify, produce, organize, store, share, use, learn and improve.
3. Perform a comparison between KM processes and RM processes.
4. Elucidate each methods and tools within the selected list.
5. Identify the RM process that could be supported by KM methods and tools based on KM processes as shown in the example below.

KM Process	KM Tools and methods	Risk Process
Identify	Environmental Scanning	Identify monitor

The first step after collecting data from the interviews was the process of data analysis. For this purpose, the researcher used a manual coding using both MS Word and MS Excel. Computer software packages for qualitative analysis were not used as it was not seemed to be needed with this sample size of interviews.

5. There is gradually developing of KM, which is slowly penetrating into the organizational strategy. The form of this developing was in gathering social networks for informally sharing experiences. For its facilitation, organizations open to suggestions environment to transform ideas into decisions.
6. On the RM facade, organizations strive to revise the SWOT analysis to align tasks to their objectives.
7. All participants agreed that suggested KM and RM models are valid and applicable.
8. Contributors were aware of all discussed methods and tools and believe that it would add value to the RM processes.
9. As per practitioners answers commonly used tools and methods are classified as per KM process. Moreover, researcher highlighted the RM processes that might use KM tools and methods, which summarized within the next tables.
 - a. Table below identified tools and methods used in Identify process as KM. These tools and methods could be useful in RM processes: identify and monitor.

Table 3. KM identify phase.

KM Process	KM Tools and methods	Risk Process
Identify	Environmental Scanning Knowledge Cafés Communities of Practice Collaborative Virtual Workspaces Expert Locator Knowledge Mapping	Identify monitor

b. Table (4) below recognized tools and methods used in KM producing stage. These tools and methods could be used in RM processes: assess, analyze and plan.

Table 4. KM methods and tools.

KM Process	KM Tools and methods	Risk Process
Produce	Brain Storming Learning and Idea Capture Communities of Practice Collaborative Virtual Workspaces Collaborative Physical Workspaces Knowledge Bases (Wikis, etc.)	Assess Analyze plan

c. Table (5) identified tools and methods used in KM organizing stage. These tools and methods could be useful in RM processes: assess, analyze and plan.

Table 5. KM methods and tools.

KM Process	KM Tools and methods	Risk Process
Organize	Knowledge Mapping Knowledge Auditing Knowledge Bases (Wikis, etc.) Expert Locator Taxonomy Collaborative Virtual Workspaces	Assess Analyze plan

d. Table (6) identified tools and methods used in KM store process. These tools and methods could be used in RM processes: store, plan and monitor.

Table 6. KM methods and tools.

KM Process	KM Tools and methods	Risk Process
Store	Building Knowledge Clusters Taxonomy Knowledge Bases (Wikis, etc.) Collaborative Virtual Workspaces Document Libraries Knowledge Portal	Store plan monitor

e. Table (7) named tools and methods used in KM share step. These tools and methods could be useful in RM processes: plan and implement.

Table 7. KM methods and tools.

KM Process	KM Tools and methods	Risk Process
Share	Communities of Practice Knowledge Cafés After Action Reviews Knowledge Portal Story Telling Peer Assist	Plan Implement

f. Table (8) specified tools and methods used in KM Identify process. These tools and methods could be help in RM processes: implementation and monitor.

Table 8. KM methods and tools.

KM Process	KM Tools and methods	Risk Process
Use	Knowledge harvesting Knowledge mapping Knowledge auditing intranet Expert Locator Advance share tool	Implement monitor

g. Table (9) identified tools and methods used in KM learn process. These tools and methods could be useful in RM learn stage.

Table 9. KM methods and tools.

KM Process	KM Tools and methods	Risk Process
Learn	Learning and Idea Capture Learning Reviews After Action Review Collaborative Virtual Workspaces Community of Practice Document Libraries leading to a Document Management System	Learn

6 CONCLUSION

In this paper, The researcher combined between academic point of view and the practitioners and experts perspective of KM and RM. The scope of the research was made clear with breakdown of the aim into defined objectives. The realization of the objectives is presented as follow:

1. The first objective was to perform a comparison study between KM and RM processes. Primary

and secondary data were analyzed and revealed that there is a high level of similarity between KM and RM processes.

2. The second objective was to identify and examine the tools and techniques of KM which can be used in RM process. Total of 20 KM methods and tools, adopted by the most successful organizations, have been recognized.
3. The third goal was to investigate whether KM methods and tools can be effectively implemented in order to enhance current RM practices. The study revealed the opinions of experts and practitioners, which is KM tools and techniques could be used to enhance RM process. Potential tools and methods for each of the RM processes have been identified and briefed in the Tables (3 to 9).

Finally, this paper concluded that organizations cannot manage their risks effectively unless they manage their knowledge. Successful integration of KM and RM can be achieved through a balanced assortment of technology, effective processes of work and adaptive mind set of people.

REFERENCES

- [1] Aghili, Shaun. (2010) Organizational risk management: successful achievement of business objectives hinges on the organization's ability to manage risk effectively. (BACK TO BASICS). Highbeam Research. [Online]. Available at: <http://www.highbeam.com/doc/1G1-229068994.html>. (Accessed: 21 August 2015).
- [2] Lottering F, Dick A. (2012). Integrating knowledge seeking into knowledge management models and frameworks. [Online]. Available at: <http://www.sajim.co.za/index.php/SAJIM/article/view/515/575>. (accessed: 21 September 2015).
- [3] Heisig, P. (2009) Harmonisation of knowledge management: Comparing 160 KM frameworks around the globe, *Journal of Knowledge Management*, vol. 13, no. 4, pp. 4–31.
- [4] Evans M, Dalkir K and Bidian C. (2014). "A Holistic View of the Knowledge Life Cycle: The Knowledge Management Cycle (KMC) Model" *The Electronic Journal of Knowledge Management Volume 12 Issue 2* (pp85–97) available online at www.ejkm.com (accessed: 21 September 2015).
- [5] Weinberger, David. (2010). *data-information-knowledge-wisdom hierarchy*. Harvard Business School Publishing. [online]. Available at: <https://hbr.org/2010/02/data-is-to-info-as-info-is-not> (Accessed: 22 September 2015).
- [6] Kothari, C.R. (2004). *Research Methodology: Methods and Techniques*. Delhi: New Age International.
- [7] Bellinger, G., Castro, D., & Mills, A. (2004). *Data, Information, Knowledge, and Wisdom*. [Online]. Available at: <http://www.systems-thinking.org/dikw/dikw.htm> (Accessed: 21 September 2015).
- [8] WebFinance. (2015). *Business Dictionary* [Online]. Available at: <http://www.businessdictionary.com/definition/knowledge.html> (Accessed: 21 September 2015).
- [9] Nonaka, I., & Takeuchi, H. (1995). *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. New York: Oxford University Press.
- [10] Choucri, N. (2007). *The Politics of Knowledge Management*. Massachusetts: Massachusetts Institute of Technology [Online]. Available at: <http://www.portal.unesco.org/education/es/files/54909/...pdf/Choucri.pdf> (Accessed: 14 August 2015).
- [11] Nicolini, D., Gherardi, S., & Yanow, D. (2003). *Knowing in Organizations: A Practice Based approach*. New York: M.E. Sharpe.
- [12] Chawla, D., & Joshi, H. (2010). 'Knowledge management practices in Indian industries—a comparative study' *Journal of Knowledge Management*, 14(5), pp. 708–725 [Online]. Available at: doi: 10.1108/13673271011074854 (Accessed: 23 September 2015).
- [13] NASA. (2008). *Knowledge Management* [Online]. Available at: <http://km.nasa.gov/whatis/index.html> (Accessed: 25 August 2015)
- [14] Lottering F, Dick A. (2012). *Integrating knowledge seeking into knowledge management models and frameworks*. [Online]. Available at: <http://www.sajim.co.za/index.php/SAJIM/article/view/515/575>. (accessed: 21 September 2015).
- [15] Evans M, Dalkir K and Bidian C. (2014). "A Holistic View of the Knowledge Life Cycle: The Knowledge Management Cycle (KMC) Model" *The Electronic Journal of Knowledge Management Volume 12 Issue 2* (pp85–97) available online at www.ejkm.com (accessed: 21 September 2015).
- [16] Young, Ronald (Editor). (2010). *Knowledge Management Tools and Techniques Manual*. Asian Productivity Organization.
- [17] PMI (2012), *A Guide to the Project Management Body of Knowledge*, 5th Ed.
- [18] Hillson, David. (2015). new concepts in project risk management. the 15th International PM Conference of the PMI Arabian Gulf Chapter held. Bahrain. [Online]. Available at: <https://www.youtube.com/watch?v=9zQ1VDS90II> (accessed: 11 September 2015).
- [19] IRM. (2002). *A Risk Management Standard* (pp. 1–16). London: The Institute of Risk Management.
- [20] Dey, P.K. (2010). 'Managing project risk using combined analytic hierarchy process and risk map' *Applied Soft Computing*, 10, pp. 990–1000 [Online]. Available at: doi:10.1016/j.asoc.2010.03.010 (Accessed: 14 August 2015).
- [21] ISO. (2009). *Risk management—Principles and guidelines*. Geneva: ISO copyright office.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Supporting knowledge society using digital repositories

Ebtesam Hussain ALZahrani

Information Science Department, King Abdul-Aziz University, Jeddah, Saudi Arabia

ABSTRACT: Generally speaking, the study being dealt with sought to give an overview of the essence of the digital repositories used in the university educational practices and to rivet the attention of those who are interested in technical education to the importance of the digital repositories in reshaping and introducing the educational content in addition to shedding light on the effective digital repositories role via the educational practices which coincide with the realm of the university e-learning. In the same vein, this study will explore in more details the earlier studies and detect the role of the digital repositories in supporting the cognitive society and how every society utilizes these digital repositories to gain knowledge and analyse it by means of SWOT analysis.

Integrated into many fundamental findings is the adoption of the analytic descriptive curriculum for more than scientific study in different universities, but the foremost of which boiled down to the fact that the institutional repositories gain much more importance, and of particular, the universities, and the research centres to the effect that the digital repositories support for knowledge were ample for its being the first source of knowledge society bolstering in spite of the low indicator of digital repository concept in most people minds to be added to many members mis usage of the digital repositories in the educational practices because of the tough challenges and impediments precluding the possibility of its usage, the most salient of which are unawareness of the functions and characteristics of digital repositories and the uphill tasks of using some digital repositories for some of the beneficiaries and the university lack of mechanism for keeping and constructing the educational repositories in the electronic libraries and the insufficient skill to select the adequate digital entities for the educational content and searching for it in the digital repositories. All these factors conspiring obstruct the digital repositories as required added to the cluster of findings through which the other recommendations were formulated.

Keywords: Digital Repositories, Academic Digital Repositories, Institutional Repositories, Mansoura University repository, Cognitive Society

1 INTRODUCTION

The most prominent feature of the 21st century is the radical transformations through which many challenges and opportunities were posed in all fields, that is, the e-learning techniques witnessed remarkable development and widespread to the effect that this development far outstretched to the development of the university educational techniques.

Moreover, the modern approaches in education technology has contributed to the emergence of new and advanced systems for teaching and learning, which had a greater effect in triggering changes and positive developments on the way students learn and on the methods and techniques adopted convey the scientific information to them and on the content and format of the school curricula in consistency with these approaches.

Among the systems which were yielded by the modern orientations of the education technology

is what was touted as “e-learning”, which depends on the employment of computer, Internet and the diverse interactive means in the education process, to be added to realizing how far knowledge is of paramount importance to the effect that it considers technology the most salient element.

And on making much ongoing headway with the electronic learning technology and in the light of the increasingly used technologies, the expectations have changed completely since the development of content by means of the educational entities produces educational sources premium quality, stressing that the educational sources offer enormous educational opportunities. In the same context, the educational entities are always maintained in reachable system format via the internet to which the term “digital repositories” is attributed.

Such being the case, reaching knowledge is a prerequisite for all sectors of society and significantly important for human development conducive to which is the free access to information, appealing

for it and relying on it as it is considered one of the best scientific communication means between researchers, students and teaching staff through the scientific product accessibility and benefit.

At this point, digital repositories are essential reference for its beneficiaries in all firms. The digital repositories are one of most cutting edge digital institutions on the Internet, and these repositories have emerged in order to ensure free access to information initiatives, the most well-known types of which are “Institutional digital repositories” which always applied in a university or in a scientific or research body. These institutional digital repositories generate the intellectual production of the scientific institution employees into digital form on the internet for free.

Furthermore, the digital universities repositories review and include all messages in terms of hardware, software, metadata elements, human efficiencies, the size of the messages’ digital repository and how to search and flip through in the repository, and also to how to search as well as the provision policy of the scientific researches close at hand for the teaching staff members and the database of the teaching staff members researches to be added to the quantities ratio of locally and internationally published researches in addition to the inclusion of the teaching staff members’ full-text researches within.

By inference, it is arguable said that the digital repositories are extremely important in general and in particular, notably in universities, that is, they contribute to maintaining and managing its intellectual assets, achieving a better global standing between universities, and upgrading the scientific research quality and educational process as a whole.

Thus far, this abstract implicative is to the core of this objective which also helps illustrate the role of digital repositories specifically in the scope of knowledge, or the cognitive society which can be shared and exchanged between the universities and research institutions and between each other and between universities, research institutions and researchers, and between researchers themselves, adding to obtaining the necessary sources of curriculum development. Accordingly the researcher reviewed the study problem pivotally as regard to the role of digital repositories in support of knowledge society with an overview of the background of previous studies to be propounded for the purpose of determining the intellectual and cognitive frames of the crux, not to mention pinpointing the study problem analytically and descriptively from the real life as forged virtually by the female researcher and reaching definite conclusion which in its turn manifests how far the digital repositories play an important role in rendering services and acting as a spur to the cognitive society.

2 RELATED WORKS

Some studies relevant to the present research, namely “the role of digital repositories in support of a cognitive society” are elicited through previous studies specialized in the field of libraries and information at the Arab level which were germane to the digital repositories and to the free access sources of information.

We will review previous studies and the most important findings in the field of digital repositories:

The first study (1) scrutinizes the identification of the digital repository of the University of Mansoura in terms of construction, content, techniques, tools and policies pursued either in the deposit or retrieval, and availability, as well as identifying the strengths in warehouse, and the weaknesses points which have to be eliminated.

This study has separated its objectives in several other subsidiaries goals as follows:

- Analytical description of the digital repository of the University of Mansoura.
- Identifying the total volume university messages in the university, and the percentage of the full text messages.
- Identifying the scientific researches volume of the teaching staff members and the ratio of full-text researches.
- Identifying the published articles size in local periodicals at Mansoura University, and the proportion of full-text articles.
- Identifying the deposit policy and availability of repository digital contents.
- Identifying how to keep the digital sources policy, and how to manage the property rights.

The most prominent findings of this study were encapsulated as follows:

The digital repository of the University of *Mansoura* represents one of the four components of the library management future system, in addition to the integrated library program. Digital repositories of the *Mansoura* university also covers: university messages repository, university teaching staff members’ research repository, and the digital repository of the scientific journals published by the university. In addition, it redresses the university messages since the stage of registration or enrolment and introducing its own metadata, added to the study plan and compiling report accordingly. The system also lets people search for messages at the bibliographic level and at the full text level. It also allows the searching and browsing process within the message.

The second (2) aims to shed the light on the importance of digital repositories through a project divided into several stages since the soci-

ety culture changes towards the free access to the launch of the repository on the internet, its assessment and evaluation starting with the first stage typified into the digital entities meaning and the digital repositories on the internet, and eventually ending with the fourteenth typified into the adoption of the proposed envisage and setting the project in motion. The most prominent findings of this study were conducted as follows:

The digital repositories meet the needs of beneficiaries on the internet, and as considering that digital repositories is somewhat unprecedented and innovative, a change in culture of the society towards the entities, digital repositories and free access to information can be wrought through three levels: - intellectual aspect (conceptual, logical, benefits) and emotional aspect (hearts—minds—enlightened personal interest) and the management aspect (Obligation or deposit—international trends).

The third study entitled “digital repositories of academic institutions and their role in educational and research process and the preparing a mechanism to setting up a digital repository of Arab Universities” aims to achieve more than other targets, of which is the analytical description of the best institutional repositories, as well as the distribution of the digital content of the international academic institutional repositories numerically, objectively and qualitatively in addition to identifying ways of providing content and digital assets for international academic institutional repositories. Standing also as an aim is to identify the software touching on the internationally used digital repositories and devising a mechanism for setting up digital repositories to manage content and digital assets in terms of keeping and retrieval.

The most prominent feature of these study findings are encapsulated as follows: All repositories included in the study made it possible to download the full text of all information sources to all users for free. Also, all repositories included in the study were keen on providing the sources retrieval option by means of browsing through sources depending on the theme or topic.

The fourth study main goals whose title is “An Evaluative Study of Some Selected Libraries in India Undergoing the Process of Digitization-2008” are: to carry out an in-depth analysis and evaluation of digital repositories at the national level and at the digital library level, performing an analysis and evaluation of digital repositories to gain access to knowledge and scientific literature as well as evaluating the digitization of different types of repositories groups such as rare books, manuscripts and newspaper articles, documentary heritage, dissertations in addition to the study of the functioning of the digitization warehouses in various projects of the digital library.

The most prominent feature this study findings are encapsulated as follows: The digital repositories initiatives as discussed in chapters 2–5 with a view to produce a huge amount of digital documents is the best way for the registration of human knowledge, ranging from rare manuscripts of the current research in literature to inferring that the digital repositories in India are the largest digitization initiative in which more than ninety organizations took part.

The fifth study namely (5) aims to explore what extent we can rely on student work in support of digital repositories, through a poll of those responsible for the management of sixty digital repositories in six countries: the United Kingdom, the United States, Canada, Ireland, Australia, and Singapore, noting that responses were received and echoed from thirty-five repositories belonging to universities, mostly from the United Kingdom.

The most prominent feature of this study findings are encapsulated as follows:

The study has demonstrated relying on student work is possible, but on certain conditions, noting that the university messages were the most turnout sources that can be incorporated in repositories to the effect that the study demonstrated the need for setting clear and guiding rules for the quality control of the digital repositories contents.

The sixth study entitled (6) aims to assess the Cornell University repository by examining its contents and how far the teaching staff members participated in comparison with three institutional repositories which used Dspace program which is commonly used by Cornell University repository.

The study also derived its results from personal interviews of the teaching staff members in the fields of science, humanities and social sciences, which reached eleven members in number aiming to know the reasons for the minimal use of the university repository.

The most prominent feature of this study findings are encapsulated as follows:-

They demonstrated that the poor contents in Cornell’s DSpace repository were the main reason for not using it in addition to the startling paucity of the teaching staff members knowledge of how to deal with the repository and the fragile protection of copyrights; definitely of the agenda particulars embedded in the repository.

3 PROBLEM FORMULATION

The importance of the study lies in the importance of digital repositories for universities, colleges and research centres for its being serving many purposes and objectives, that is, they positively contribute to the upgrading of scientific and educational research quality in general through the help of

these educational institutions in maintaining and keeping its intellectual assets, and its management by means of linkage between the content of these different types of researches and between the societies which applies this strategy on the ground. The importance of institutional digital repositories lies in serving the following sectors:-

- Universities and research institutions
- Scientific Research
- Students and teaching staff
- Society

As long as scurrying towards the importance of the role played by the academic repositories in academia media is a prerequisite to the inclusion of all whoever is concerned in the domain given the scarcity of case studies of repositories in universities, there was a persistent need to be briefed on the reporting case of the digital warehouse in terms of its importance and its salient role in the cognitive society.

The idea of this study stems from the interest of the researcher for employing technical innovations in university education effectively, in addition to the importance of the digital repositories topic nowadays in the higher education system. By inference, the current study scrutinizes the importance of the digital repositories role in rendering cognitive society service, bearing in mind what can be discerned by the researcher; definitely the mediocre technical skills of a lot of users.

Such being the case and immediately on the researcher review of the theoretical literature and of previous studies, she was doomed to feel the scarcity of research and study which dealt with the use of digital educational repositories for rendering service to the desired goals, despite their importance and educational attitudes in university education.

Based on the above, the problem of the research is latent in the fact that despite the importance of digital repositories and for being a necessity for educational institutions, there was still a lack of interest in highlighting its role in cognitive society service in particular. The questions took the form of following:-

- What is the role of digital repositories as regards rendering service to the cognitive society?
- How far does the beneficiary use and if his needs of the digital repositories are met?
- To what extent have the digital repositories contributed to the development of universities?

4 THE USED RESEARCH APPROACH AND METHODS OF DATA COLLECTION

The researcher used the descriptive analytical method because it fits with the nature of this study, whereas it hinges on the study of the phe-

nomenon in actuality and thrust into focus its accurate description and expresses it qualitatively and quantitatively for studies which include digital repositories in general and in detail of the various universities and free access, and coursing through the discussion of the international best academic institutions and programs used to create them, in addition to the nature of digital content and assets of these international academic institutions, and eventually the reflection of the intellectual production.

4.1 *The study subjects*

The researcher is charting her strategy to select the study subject which targets three different studies picked through interviews conducted with the university digital repository superiors underscoring its entity as a system in terms of its construction and development and with repository superiors on the librarian and messages digitized perspective as well as obtaining a copy of local university journals, and the scientific researches of the teaching staff, represented in the central library at the university.

4.2 *Study tools and objective*

In order for the objectives of the study to be attained, the researcher examined the previous studies on digital repositories, analysed them, and gleaned in-depth data of relevance to preparing the various sources of information and the potentials of the research, added to flipping through and pondering on the policies used and other aspects covered by the study through the interview and the Internet. The authors use SWOT analysis tool to be aware about the role of digital repositories in supporting the cognitive society.

5 STUDY FINDINGS

In the light of the previous studies and propelled by the importance of the role played by digital repositories, the study aimed to determine the role magnitude of digital repositories on society service, the degree to which the beneficiary uses these repositories if his needs are met to the core of digital repositories (see Table 1).

5.1 *The first question results*

The first question was entitled “To know of the extent to which the role of digital repositories affects the cognitive society service”?

To answer that question, we can detect the volume of the positive return for the use of digital repositories and its importance in rendering services to cognitive society whether was specifically

Table 1. The volume of the positive return for the use of digital repositories.

Contributor	Institution	User
Accessibility	The scientific subjects churned out by the university are available in one place serving to be a reflection of the intellectual achievements	The subjects are accessible in the digital repositories through the search engines
Increasing rate of reference citations	The documents which reflect the university institutional history whether were scientific or non-scientific to the effect that they are actually kept for later use .	No subscription or entry fees
Malleability(Easy to use)	Highlighting the type of the intellectual capital of the institution and making use of the investments in information systems .	Non-commercial literature (Grey Literature) are rare subjects to find in its traditional form
Organization		Dilating on knowledge fields

Table 2. Meets the needs of digital repositories users online.

	Means	Purpose	Responsibility	Targeted society	Recurrence
1	The most commonly circulated questions	Answers for repository questions	Administrator	The whole world	When needed
2	Meeting with Beneficiaries	Communicating with beneficiaries	Administrator	Beneficiaries	When needed
3	E-Bulletin	Broadcasting information about the repository	Administrator	Beneficiaries	Continuous

for the contributor or the institution or even for the user.

5.2 The second question results

The second question, which was entitled as “The extent to the beneficiary uses digital repositories and if his needs of them are met?”

To answer this question, it is doubtless that the launch of the repository does not mean that its planning came to an end, but you must develop a plan to meet the needs of beneficiaries from the repository on the long run. Therefore, some requirements have been proposed as shown in Table 2

5.3 The third question results

The third question was entitled as “the extent to which digital repositories contributed to the development of universities”?

The study reached definite conclusions about the advantages and benefits of digital repositories in universities and its impact on development, such as:

- Ease of the information exchange between researchers of different specializations

- Communication between peoples, the cultural and informational exchange as achieved by free access

6 CONCLUSIONS AND RECOMMENDATIONS

The increasing number of digital repositories rate was tantamount to a repository emergence a day in the world to the extent that the prospects for finding a serious research institution which has no a digital repository at the end of the decade were unlikely and preposterous. Therefore, myriad of huge investment programs were investment-oriented in some countries for the purpose of establishing a network among these repositories to facilitate the scientific communication and to increase the productivity of the research community.

To the core of what can be deduced from the research and to what can be learned is the fact that importance is attached to the digital repositories especially those in universities and research centres because they provide the potential to save the digital content of the institution associates. No less important, the content can be broadcasted and managed to the effect that the digital repositories is conducive to exchanging information and expertise at the local, regional and international level and we now know full well their contribution to the curricula development processes.

6.1 Conclusions

- The researcher hypothesizes that the results shown in the table above are consistent with some studies which revolve around how important the role of the digital repositories and around the strengths central to its usage and challenges as well the foibles which constrain the usage of these repositories.
- Also, the studies and results based on them made it clear that there are many obstacles faced by the learners on making use of the digital depositories, but these studies elucidated as well as the usefulness and importance of these repositories which are central to the service and cognitive society support.
- The foregoing may be due to the novelty of the educational digital digital repositories and the unclear idea among workers in educational institutions, as well as the difficulty of its design and its production according to high quality standards. In the same context, a lot of researchers asserted the importance of being advertent to the high education techniques for its role which is apparently latent in reshaping universities and rapid change in full throttle of technology and the globalization of institutions and other future challenges.
- In general, results of the study cropped up with consensus to the importance of the use of dig-

Table 3. SWOT analysis.

Strengths	Foibles	Threats	Opportunities
Providing opportunities for exchanging knowledge and information A mechanism is wrought to control quality in repository	Changing approaches for saving and keeping Digital repositories need successes and quick gains The digital repository might fail if the institution stopped financing it . Needs support from up to down and vice versa	Inflexible educational process Need technical support as necessary as possible The necessity for training those who are in charge of dealing with it	Providing better access Points for the information sources Stressing the high-calibre of the digital repositories.

ital repositories and their very important role to serve the cognitive society and the importance of adopting educational practices despite low clarity indicator of the digital entities repositories concept for most in addition to the beneficiaries and members misuse of digital repositories.

- Thus, the digital repositories act as modern informative institution which have objectives, disciplines and its policies are like any informative institution, whether traditional or digital.
- To sum up, they are like collaborative work online space for collecting and maintaining the scientific output of the academic institutions and research centres.

6.2 Recommendations

In the light of previous results, we can recommend the following important points:-

- Expanding on establishing and developing the digital educational repositories in the light of quality standards.
- Awareness-raising of paying attention to the educational digital repositories which store content and provide a large number of links and instigate universities into the pursuit of establishing cooperative framework.
- The establishment of centres for designing and producing educational digital repositories content in various fields and to be supported financially.
- Backing researches of relevance to the free access movement upgrade it to a competitive level approaching the developed countries in this field.

- Preparing a list or directory of digital repositories to help users in universities employ this new technology.
- Conducting a study on the researcher's role in supporting the digital repositories whether in terms of establishment or reports.

In the same vein, some topics are suggested for further prospective study as follows:

- Designing and constructing a digital repository for the research projects in the field of education learning.
- Conducting a research study on the impediments of the digital entities repositories in the teaching practices in universities.
- Conducting a study on evaluating the Arab educational digital repositories.

REFERENCES

- [1] Mansoura University- Digital repository of Mansoura University: A case study of The digital repository in the future System for library management.
- [2] "Majmaah University - [Institutional Digital Repository for IDRMU)2011).
- [3] Ahmed Abada – Tanta The Digital repositories of the academic institutions and its role in the research and educational process, adding the preparedness of a mechanism for setting up a digital repository for Arab universities-Egypt).
- [4] Omar-Eman Fawzy (The technology of libraries and information – Third Generation techniques and its entries in the libraries and information society) – Alqadera: The network of the information and libraries specialist (net www.motlaqa.libriannet)
- [5] Pienaar, M. v. (2008). South African Repositories: Bridging Knowledge Divides. *Ariadne* (55), <http://www.ariadne.ac.uk/issue55/vandeventer-pienaar/>.
- [6] Connolly, P. M. (2007). Institutional repositories: Evaluating the reasons for non-use of Cornell University's installation of DSpace. *D-Lib Magazine*, 13 (3/4).
- [7] McKnight, M. P. (2007). Is there a role for research students in an institutional repository? Some repository managers' views. *Journal of Librarianship and Information Science*, 39, 153–161.
- [8] An Evaluative Study of Some Selected Libraries in India Undergoing the Process of Digitization 2008.
- [9] ODLIS — Online Dictionary for Library and Information Science (2011) Retrieved January 21 2011, from http://lu.com/odlis/odlis_i.cfm.

Impact of information resources on decision-making process in different enterprises

Hessa Mouner Albogami

Department of Information Science, King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT: There are many factors involved in decision making process for any enterprise and each of the core skills has the potential to impact effective decision making. In an ideal world decisions would be made objectively, with a full set of evidence, an endless bank of resources, no time pressures, minimal interruptions, decision support tools to hand and plenty of energy to handle any decision making situation at any time of the day. So, the major objective of this paper is to shed the light and discuss the relationship between information resources and decision making process. This study was conducted in Saudi Arabian Airlines organization in Jeddah, the survey method was used in this study, and the data collection tool is a questionnaire, which was distributed to members of the sample who reached the 23 employee of the Foundation Saudi Arabian Airlines in Jeddah. The main finding show that providing the various information resources in a timely and effective use contribute significantly to support decision-making within organizations. Also, controlling the information flow will help consumers better match their preferences, have better memory and knowledge about the domain they are examining and be more confident in their judgments. One of the main recommendations of this article need to educate managers makers of administrative decisions in the Saudi Arabian Airlines the importance of a library or information center in the organization and the need for information specialists well qualified institution in the Saudi Arabian Airlines as well as like institutions.

Keywords: MIS, HR, CIMA, HRIS and Saudi Airlines

1 INTRODUCTION

Management information systems is no doubt an efficacious tool for modern business practitioners; its role in decision-making cannot be over emphasized because effective decision-making is ultimately a function of accurate, timely, relevant, complete and economical information, which MIS produces. Information resources have become of high demand in today organizations and MIS is the only option for the satisfaction of such demands. Entrepreneurial spirit and business judgment (that human ability to weigh intangibles and ambiguity) will always be important in decision making [1]. But the risks of personal bias, repeating past mistakes, acting on guesses or following hunches unnecessarily, can be limited if a culture of evidence based decision making is fostered. Providing evidence in the form of financial and management information has long been the basis for accountants' role in the decision making process. Supporting the strategic planning process and providing the metrics and analysis to support evidence based decision making are important. But these will no longer suffice.

Historically, human resource information has largely been seen as a necessary tool in the hiring, administration, and, ultimately, separation of employees. Over the years, these processes have not changed dramatically, but the way the information is gathered and stored has. Going back to a time prior to the proliferation of technology in the workplace, an employee would submit a paper application to a prospective employer. There would be general information about the employee, including name, sex, age, social security number, employee's address, education, marital status (in some cases), employment history, and so on. This information would be stored in a folder for the HR department to access as necessary [2]. Once the employee was hired and placed on the payroll, the application could be used to provide some information to the finance department for pay purposes, while other information could be used internally by the HR department to track hiring practices and recruitment. Over time, additional information about the employee would be placed in the folder, including benefits, performance reviews, promotions, discipline, and training. Generally, employees were responsible for updating their own records, while the employer was the

primary custodian of all such employee related information. Most employers viewed this information as necessary but not particularly valuable from a strategic viewpoint. After all, each employee came into the company at a different time, progressed through his or her career at a different pace, and left the company for differing reasons, under differing terms, and at different times.

Effective decisions are those that achieve impact. An effective decision making process spans from how strategic decisions are informed and considered, through how performance and risk are assessed and managed, to how routine operational decisions are guided, made and governed so the intended impact is actually achieved. Management accountants who can combine financial expertise with business understanding have the potential to support decision making in a wide range of roles throughout this process. Training as a management accountant combines accounting and business disciplines so it helps to develop accountants who can support decision making [3]. As CIMA's focus is on qualifying accountants who meet employers' needs rather than for public practice, the CIMA syllabus reflects employers' expanding requirements. The CIMA qualification is for people with ambition that is broader than becoming an accountant. Nowadays, employers don't just want accountants with the technical skills to produce accounts. They want accountants who can apply financial expertise in support of the business and contribute to leadership. The domain of management accountancy has expanded in line with this general trend in employers' emphasis from technical skills and financial reporting to management skills and decision support through to impact. The human resources information systems can benefit any organizations in various directions like [3]:-

- *Improving productivity*

One of the most important of all HRIS benefits relates to the ability of the software program to improve productivity of human resources employees. These systems are highly detailed, and they are designed to enhance and speed up the efforts of HR employees in a number of ways. For example, they can assist with the recruitment process by simplifying the efforts associated with collecting resumes, reviewing candidate information and more.

HRIS systems can also be used to improve productivity related to financial management through payroll processing tasks and benefits administration. These and other related tasks may require numerous hours of manpower each week, but the time and effort required to complete them can be drastically reduced when some of the tasks are automated through an

HRIS system. Tasks that may have required the support of numerous employees and that may have required many hours of labor may become tasks that can be completed quickly and easily with the software program.

- *Reducing errors*

Many HR tasks are highly regulated, and because of this, even a minor error on the part of a human resources employee could result in considerable legal issues and even financial loss for the company. For example, when resumes are not reviewed in a fair and just manner during the hiring process, a lawsuit may ensue. Another example involves a seemingly small accounting error with payroll processing but a small payroll error could yield considerable financial expense for the company. When considering HRIS benefits for your organization, the ability to reduce these and other related errors associated with human oversight or other factors can be considerable. Furthermore, additional HRIS benefits relate to compliance issues. Some software programs are designed to review compliance with specific rules and regulations—this makes it easier to ensure that your company is in compliance with these laws and regulations.

- *Performing analyses*

Additional HRIS benefits relate to performing analyses and reviewing metrics associated with various aspects of the organization. For example, the human resources department is responsible for analyzing hiring costs and for calculating the turnover rate in different departments. This information can be difficult to properly and accurately determine, but the accuracy of these calculations is imperative. The results of these calculations may be used to make important business decisions and to develop strategies to move the organization forward along a successful path.

HRIS benefits associated with analytical tools give HR employees the ability to perform these calculations with speed by collecting the data needed within a short period of time and by analyzing all of the data in a concise and effective manner. Some software programs are designed to create professional reports on metrics and analysis as an additional benefit to the human resources professionals. There are many HRIS benefits that companies can begin using effectively once the human resources information system has been implemented in the organization. There are several different types of HRIS systems that can be purchased and implemented, and each may offer different features and functions. Companies should review the different options available carefully in order to find the right program for their needs and budget.

2 RELATED WORKS

The information needs of modern organizations have become quite enormous and challenging to the extent that every organization needs to pay great attention to how information is gathered, stored, disseminated and utilized. This situation has arisen because of factors such as increased organizational size, expanded operational scope, competitive influence and overall environmental vagaries. Also, today's organizations require tools to support quicker and automated decisions, as well as ways to minimize uncertainty; only an effective management information system can ameliorate this challenge. In the following paragraphs, we'll provide some of our literature reviewing covering some works in the benefits of information resources rules in decision making process. In [4], the study examined the extent of the use of the faculty of the College of Computer and Information Sciences members at the University of Imam Muhammad bin Saud Islamic University in Riyadh sources the electronic information systems. The main finding of this work has revealed that all faculty in the College of Computer and Information Sciences members of Imam Muhammad bin Saud Islamic University in Riyadh (study population) use electronic information sources, where the percentage of that 100%, as the results of the study showed that access to information faster. Then the novelty of information is the main reasons and the reasons for the use of faculty members to electronic information sources.

The increased use of web technology to deliver HR will leave HR specialists more time for strategic decision making and that outsourcing of people-management activities will liberate HR specialists to perform more strategic activities [5]. According to [6] as one of the strategic partners, the HR manager derives benefit from IHRS, to disseminate and execute the strategy within the organization. These systems enable employees to manage much of their own HR administrative work [7]. They can take care of many routine transactions whenever they wish, because automated systems don't keep office hours. In addition to their former operational role, HR professionals can also act as a competency manager by arranging the right people to the right positions in the right time with their new strategic architecture role [8]. HRIS is thought to contribute to overall business performance by fulfilling or at least supporting the tasks of data storage and retrieval, of serving as primary administrative support tools, of reporting and statistics as well as of program monitoring [9]. HRIS plays an important role for any organization to effectively manage its human assets. Many organizations have adopted HRIS to assist their daily human resources operations. HRIS must align and satisfy the needs of the organization and its users in order to be successful [10].

In study [11] entitled "The use of faculty members at the University of Kuwait to sources of digital information, the aim of the study to identify the extent of the use of teaching Kuwait University faculty members to sources of digital information. It use descriptive approach as an appropriate method to achieve the top scorer in this study through a sample of 180 faculty at Kuwait University, members of the test was the most important results are the majority of the teaching staff in the scientific and literary colleges members agreed that their use of sources of digital information is a necessity and feel the importance of their presence. Skills especially for the use of sources of digital information are difficult for faculty members earned. In order to overcome these obstacles, the study advises on promoting further courses for faculty members that help familiarize them with the evolution of their research skills in the sources of digital information.

3 PROBLEM STATEMENT AND METHODOLOGY OF SOLUTION

There is no doubt that both information systems and organizations influence one another. Information systems are built by managers to serve the interests of the business firm. At the same time, the organization must be aware of and open to the influences of information systems to benefit from new technologies. The interaction between information technology and organizations is complex and is influenced by many mediating factors, including the organization's structure, business processes, politics, culture, surrounding environment, and management decisions. So, we shall need to understand how information systems can change social and work life in your firm. We shall not be able to design new systems successfully or understand existing systems without understanding our own business organization. So, this paper tries to reply the important questions given by: what about the effectiveness of information systems resources on decision making process in Saudi Airlines Company in Saudi Arabia.

This article was using research surveys method as it is capable of obtaining information from large samples of the population. It is also well suited to gathering demographic data that describe the composition of the sample [12]. Surveys are inclusive in the types and number of variables that can be studied, require minimal investment to develop and administer, and are relatively easy for making generalizations [13]. Surveys can also elicit information about attitudes that are otherwise difficult to measure using observational techniques [12]. It is important to note, however, that surveys only provide estimates for the true population, not exact measurements [14].

4 LIMITATIONS TO DECISION MAKING

Given the distinctions made concerning decisions, it is also important to look at what organizational factors limit the authority of individuals to make genuine decisions, including human resource decisions [15]:

- Decisions are limited by organizational position. The concept of decision-making is restricted by the broader notions of authority and responsibility. That is, the authority to make a decision is directly associated with the responsibilities assigned. Often a manager has the right to make hiring decisions within her own department, because the effectiveness of the department is her responsibility; but the manager seldom has the power to make hiring decisions for other departments because the activities of other departments are not part of her responsibilities.
- Decisions are limited by regulations. Each employee, no matter how high in the bureaucratic hierarchy, is governed by laws, rules, and policies that diminish the extent of the employee's power. Even a library director lives very dangerously, if she violates accepted policies and procedures or contravenes civil rights laws or an employee's right to due process or privacy.
- Decisions are limited by the responsibilities of others whose function may be interdependent or even competing, especially when limited organizational resources are involved. For example, a department head may decide to order replacement materials for her collection (a decision well within the purview of a department head), while the head of another department may decide to purchase additional databases. If there are limited resources, the decision of one may be limited by the decision of the other.
- Decisions are limited by political and social relationships. Although a decision may fall within the formal purview of a particular position, the individual may lack the confidence of superiors, thus effectively nullifying decision-making authority. Similarly, the decision-making authority of even the most competent manager may be limited by a director who believes that all decisions should be made by him or her.

5 DECISION MAKING UNDER UNCERTAINTY

At times a decision maker cannot assess the probability of occurrence for the various states of nature. Uncertainty occurs when there exist several (i.e., more than one) future states of nature but the probabilities of each of these states occurring are not known. In such situations the decision maker

can choose among several possible approaches for making the decision. A different kind of logic is used here, based on attitudes toward risk [16]. Different approaches to decision making under uncertainty include the following:

- The optimistic decision maker may choose the alternative that offers the highest possible outcome (the "maximax" solution);
- The pessimist decision maker may choose the alternative whose worst outcome is "least bad" (the "maximin" solution);
- The third decision maker may choose a position somewhere between optimism and pessimism ("Hurwicz" approach);
- Another decision maker may simply assume that all states of nature are equally likely (the so-called "principle of insufficient reason"), set all values equal to $1.0/n$, and maximize expected value based on that assumption;
- The fifth decision maker may choose the alternative that has the smallest difference between the best and worst outcomes (the "minimax regret" solution). Regret here is understood as proportional to the difference between what we actually get, and the better position that we could have received if a different course of action had been chosen. Regret is sometimes also called "opportunity loss." The minimax regret rule captures the behavior of individuals who spend their post decision time regretting their choices.

6 NUMERICAL RESULTS

All modern organizations have certain characteristics. They are bureaucracies with clear-cut divisions of labor and specialization. Organizations arrange specialists in a hierarchy of authority in which everyone is accountable to someone and authority is limited to specific actions governed by abstract rules or procedures. These rules create a system of impartial and universal decision making. Organizations try to hire and promote employees on the basis of technical qualifications and professionalism (not personal connections). In the following paragraph and table, we'll present the importance of using information to make management decisions in well pattern form as shown in Table 1:

It is seen from the table 1 that 60.9% of the study sample using the information in the daily decision-making routine to the same degree that they use in making strategic decisions. With regards to the level of use of sources of information, preferably used to obtain information to support managerial decision-making process as shown in table 2 that illustrate the Frequencies, percentages illustrates the level of use of sources of information, preferably

used to obtain information to support managerial decision-making process

From table 2, it is clear the following important points:-

- 30.4% of the study sample is always in need of personal contacts to obtain information to support the decisions they make the administrative process, 17.4% of respondents often need personal connections.
- 34.8% of the study sample always need to co-workers and advisers to obtain information to support the decisions they make the administrative process, 21.7% of respondents often need to co-workers and advisers to obtain information to support the decisions they make the administrative process.

Table 1. Illustration of how Administrative decisions makers use the information.

Expression	Repetition	Percentage
The use of information in daily decision-making routine than I use them in making strategic decisions	1	4.3%
Use the information in making strategic decisions than I use them in daily decision-making routine	8	34.8%
The use of information in daily decision-making routine equally used in making strategic decisions	14	60.9%
Summation	23	100%

- 30.4% of the study sample rarely what they need for the library organization or company for information supports they make administrative decisions process, 30.4% of respondents also do not need at all to the Library of the institution or company for information supports they make administrative decisions process.
- 47.8% of the study sample did not need at all for the information services companies paid the price to get the information to support managerial decision-making process, 17.4% of the company need information services always.
- 30.4% of the study sample is always in need of the Internet to get information to support managerial decision-making process, 21.7% of respondents often need of the Internet.
- 34.8% of the study sample did not need at all to other libraries outside the organization to get the information to support managerial decision-making process, 17.4% rarely need other libraries outside the organization to get the information to support the decisions they make the administrative process.
- 21.7% of the study sample often needs government and official publications to obtain information to support managerial decision-making process, 21.7% always in need of government and official publications to obtain information to support managerial decision-making process.

7 CONCLUSION AND RECOMMENDATIONS

Essentially, this article identified that information resources connotes those outfit/media through

Table 2. Frequencies and percentages of the level of use of sources of information, preferably used to obtain information to support managerial decision-making process.

Information resources	Usually need		Often need		Sometimes need		Rarely need		Don't need at all	
	%	T	%	T	%	T	%	T	%	T
Personal communications	30.4	7	17.4	4	17.4	4	26.1	6	8.7	2
Co-workers and advisers	34.8	8	21.7	5	26.1	6	17.4	4	0.0	0
Enterprise library or company	17.4	4	4.3	1	17.4	4	30.4	7	30.4	7
Information services company paid the price	17.4	4	0.0	0	17.4	4	17.4	4	47.8	11
Internet	30.4	7	21.7	5	34.8	8	13.0	3	0.0	0
Other libraries outside the institution	8.7	2	13.0	3	26.1	6	17.4	4	34.8	8
Government and official publications	21.7	5	21.7	5	30.4	7	13.0	3	13.0	3
Newspapers and magazines	13.0	3	4.3	1	17.4	4	43.5	10	21.7	5
Radio and television	4.3	1	4.3	1	47.8	11	34.8	8	8.7	2
Professional and specialized courses	34.8	8	0.0	0	30.4	7	30.4	7	4.3	1
Electronic databases	39.1	9	21.7	5	26.1	6	8.7	2	4.3	1
Scientific and professional conferences and meetings	21.7	5	4.3	1	43.5	10	26.1	6	4.3	1
Publications businesses and competitors	17.4	4	17.4	4	34.8	8	8.7	2	21.7	5
Institution or company files	34.8	8	8.7	2	34.8	8	8.7	2	13.0	3

which individuals and organizations obtain relevant information that will be adjudged as an asset to the corporate existence and survival of oneself and the organization as a whole. These outfit/media included people/specialists, information technology, textual information and electronic databases and other electronic outfit such as websites, blogs etc. It is hoped that we shall use the knowledge acquired in this unit to appreciate the roles of information resources in the development of our organizations and help in decision making process as well as solving major problems in the organization [17]. We concluded that the key challenge of information resource management is the mismatch between the information provider and information users, which can be summarized as three types:

- The important information is provided but is not important for users, and therefore the provided information is of no use to information users;
- The important information required by information users is not important for information providers, and therefore it is not available for information users;
- Although the provided information is important and exists, it may not be understandable for information users because the inconsistent or mismatched description and definition of information between the information provider and users.

The author advices with some important recommendations to properly use the information systems resources as follows:-

- The need to educate managers administrative decisions makers in Saudi Arabian Airlines importance of a library or information center in the organization, and used to take important decisions, and make them aware of the importance of using modern information sources in managerial decision-making process.
- The need for a specialist information well qualified in the Foundation Saudi Arabian Airlines In-like institutions, where the presence of information specialists will help decision makers in the organization to access information quickly and effectively the largest, and will they solve the problems of access to information in the organization, such as the lack of adequate time to gather information, as the specialist can assume this task of behalf of the decision-maker.
- Requires the activation of the administrative decision-making process in the light of the strategic interest in the sources of the multiple and diverse information and surveyed in an enterprise lines of information, in addition to the assignment of specialists in the field of information, a goal President Find strategic information that serve the adminis-

trative decision, commensurate with the size and type of activity Foundation Saudi airlines.

REFERENCES

- [1] http://www.cimaglobal.com/Documents/Thought_leadership_docs/cid_execrep_unlocking_business_intelligence_Oct09.pdf
- [2] <https://www1.villanova.edu/content/dam/villanova/VSB/assets/maggitiresearch/Human%20Resource%20Information%20Systems%20Administrative%20and%20Strategic%20Advantages.pdf>
- [3] <http://www.hrpayrollsystems.net/hris-benefits/>
- [4] Turkish Kazim Obeis. (2010). Management information systems and its importance in decision-making. University of Babylon magazine.
- [5] Kulik, C. T., & Perry, E. L. (2008) When less is more: The effect of devolution on HR's strategic Role and Donstrued Image, Human Resource Management, 47(3), pp. 541–558.
- [6] Ulrich, D., Brockbank, W., Johnson, D., Sandholtz, K., & Younger, J. (2009). *İK Yetkinlikleri*. (Nazlı Şahinbaş Köksal, Trans.) Turkey/Istanbul: Humanist Press. (Original Work—HR Competencies published 2008).
- [7] http://issbs.si/press/ISBN/978-961-6813-10-5/papers/ML12_029.pdf
- [8] Gürol, Y., Wolff, A., & Ertemsir Berkin, E. (2010). E-HRM in Turkey: A case study. In I. Lee (Ed.), *Encyclopedia of E-Business Development and Management in the Global Economy*, pp. 530–540.
- [9] Ostermann, H., Staudinger, B., & Staudinger, R. (2009). Benchmarking human resource information systems. In T. Coronas & M. Oliva (Ed.), *Encyclopedia of Human Resources Information Systems: Challenges in E-HRM* (pp. 92–101). Hershey, PA: IGI Global.
- [10] Noor, M. M., & Razali, R. (2011). Human resources information systems (HRIS) for military domain-a conceptual framework, International Conference on Electrical Engineering and Informatics, 17–19 July, 2011, Indonesia.
- [11] Suet Abdulaziz Fayez sawmill Zafiri. (27 to 25 March 2013). The use of faculty members at the University of Kuwait to digital information sources in Kuwait. Local Second Conference of e-Learning, pp. 1–13.
- [12] McIntyre, L. J. (1999). *The practical skeptic: Core concepts in sociology*. Mountain View, CA: Mayfield Publishing.
- [13] Bell, S. (1996). *Learning with information systems: Learning cycles in information systems development*. New York: Routledge.
- [14] Salant, P., & Dillman, D. A. (1994). *How to conduct your own survey*. New York: John Wiley and sons.
- [15] https://www.ideals.illinois.edu/bitstream/handle/2142/4014/Rubin_OPI98199.pdf?sequence=2
- [16] <http://economia.unipr.it/DOCENTI/CILLONI/docs/files/Lesson%2004%20Reading%20facoltativa.pdf>
- [17] <http://nou.edu.ng/uploads/fileuploads/cIO49hsqEf1438349717.pdf>

The reality of using various information resources to support the managerial decision-making process: Survey study on Saudi Arabian airlines enterprise in Jeddah City

Hessa Mouner Albogami

Information Science Department, King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT: Enterprise systems have been widely sold on the basis that they reduce costs through process efficiency and enhance decision making by providing accurate and timely enterprise wide information. In the same time, the success of decision-making, which is the heart of administrative process, is highly dependent partly on available information, and partly on the functions that are used in the management process. This paper aimed to recognize the reality of the use of members of management resources in the private sector institutions to assist in decision-making and problem-solving as well as searching for the most important sources of information that institutions use in their work. The most important findings of the article that author reach is that the majority of respondents in terms accounted to 60.9% use the information resources in making the routine rules is with the same degree of daily decisions that they use in making strategic decisions. The more sources of information that were used exist in Saudi Arabian Airlines enterprise in Jeddah assists in making decisions are: personal contacts, consultation and co-workers, specialized journals, and electronic databases, files institution while the less commonly used sources of information services sources were the library whether inside or outside the organization and car magazines.

Keywords: Data, Information, Internal and External sources of information, ERP, MIS, Managerial Decision-Making Process

1 INTRODUCTION

The human should convey what he learned from the information and experience of others and kept for his children and for future generations, so they were invented what is known as the external memory, and is the media that rights record by information external memory, and media such media evolved over time, the man was an old logging information on the walls of caves and then panels of clay, and papyrus in the Pharaonic era and others. So, he developed media like paper, spread paper use significantly inventing the printer, and at the end of the last century digital sources emerged were widespread deployment dramatically in recent years with the widespread use of network Internet [1].

Nowadays, we are living in a time of great change and working in an Information Age. Managers have to assimilate masses of data, convert that data into information, form conclusions about that information and make decisions leading to the achievement of business objectives. For an organization, information is as important resource as money, machinery and manpower. It is essential for the survival of the enterprise. To achieve

the major objectives of any enterprise; they should working based on the new technologies based on computer like management information systems and other similar tools to assist in decision making and solving a lot of existing problems that cannot be solved using traditional techniques.

Management Information System (MIS) is an organized, automated, and diverse information system that gathers, stores, processes, and distributes data associated with different departments of the organization. This data is processed in various forms, such as graphs, diagrams, charts, and reports to generate accurate, relevant and valuable information for the management. This information is further communicated to the various departments to be used for decision-making and business management. MIS system provides central storage of all the business information. There are various types of MIS systems which are used to gain better understanding of the market and enterprise. MIS is used across all levels in an organization. For example, MIS provides vital information at senior levels to help make strategic decisions. At other levels, MIS observes an organization's activities and distributes information to everyone in the

organization and customers [4–6]. MIS is very important for every organization because it not only collects and manages information, but also represents it in various formats useful for the management to make important organizational decisions.

On other hand, the sources of information has a great importance in various institutions with the help of administrators in decision-making and other administrative processes, has resulted in the large size of modern administrative institutions, and the complexity of its activities and its business to increase their need for and means of fixed and methods, to gather information, analyze, and classify, and save them to be close and ready to decision-making positions. So, witness the business world today has a great interest in actively information and the introduction of modern information technologies to these institutions, given that the information has become considered an important resource for institutions of various kinds, and there is no doubt that the information is a key source in the decision-making, given the current challenges in this century where the presence of regional and global competition between institutions has increased the importance of information, and is one of the important decision-making process and administrative processes information of great importance in this process [1].

2 LITERATURE REVIEW

Information systems can be classified by the specific organizational function they serve as well as by organizational level. In the following section we have described some of related works concerning the various sources of information systems with its importance in different applications associated with supporting each of the major business functions. The sales and marketing function is responsible for selling the organization's products or services. For example, marketing is concerned with recognizing the customers for the firm's goods or services, determining the customers' needs and advertising and promoting these goods and services. Sales are worried with contacting clients, selling the products and services, taking orders, and following up on sales.

In [1] entitled: "The role of information systems in decision-making in government institutions: a field study of public institutions in the governorate of Irbid", the author assume a relationship between the methods of collecting information and the speed of managerial decision-making, and the existence of a relationship between adequate information and the effectiveness of decision-making administrative, has been

using a private questionnaire in order to identify the reality of these hypotheses was reached to several conclusions including: the information systems and technologies having an active role in decision-making in the government of the province of Irbid (**Jordan country**) institutions process, and that these technologies, especially modern ones and computerized having an important role in the speed of access to information. The researcher came out in the end a set of recommendations concerning the attention to the design of information systems for each department or institution in the province before the lesson to accelerate the project of e-government with the benefit of Arab and international experiences in this field.

In the study of [2] entitled "The management information and its importance systems in decision-making," the most important findings of the study: in order to save administrative information and the face of the vast amount of information must be saved in a variety of systems, different management information systems depending on the nature of the work of the organization or facility, and that information systems are considered systems to support management decisions and these systems can be classified according to the administrative functions to which they relate, as the researcher investigated that administrative information systems provide all the information you need in different departments to exercise administrative process and facilitate statistical measure results as well as knowing their causes. They suggest at the end of the research work to develop networking systems to suit the update information according to scientific development and the motorcade it.

Thirdly the work done by [3] aimed to review the importance of strategic information to the various departments. To check the awareness of officials of the Kuwaiti companies relating, the authors conducted their study to conduct a questionnaire included 347 charge as shown: that environmental scanning term and system information precautionary are most commonly used in Kuwaiti companies to refer to the process of gathering information, that knowledge of company officials in medium terms, strategic information collected a few, there is a close relationship between knowledge and knowledge of environmental scanning partial information and strategic information. On the one hand there is a relationship between knowledge and behavior of partial information collected in which there are no differences in the behavior of the compilation of information between big and small companies. The study concluded that the need to rely and the conduct of strategic information and

the development of information systems to assist in decision-making.

3 PROBLEM STATEMENT

The researcher is trying through this article to identify the sources of information available to the Saudi Arabian Airlines Corporation in Jeddah and methods of access to use and support the decision-making process as well as facilitate the daily work by Saudi Arabian Airlines Foundation staff. As the great importance of information in decision making and performance management processes efficiently and effectively the largest, in terms of information it has become these days is considered an important resource of ERP, due to the rapid changes and developments in the business world and modern techniques.

4 RESEARCH METHDOLOGY

The author use the descriptive research approach which is not fit neatly into the definition of either quantitative or qualitative research methodologies, but instead it can utilize elements of both, often within the same study. The term descriptive research refers to the type of research question, design, and data analysis that will be applied to a given topic. Descriptive statistics tell what is, while inferential statistics try to determine cause and effect [7].

The used research methodology can be either quantitative or qualitative. It can involve collections of quantitative information that can be tabulated along a continuum in numerical form, such as scores on a test or the number of times a person chooses to use a-certain feature of a multimedia program, or it can describe categories of information such as gender or patterns of interaction when using technology in a group situation. Descriptive research involves gathering data that describe events and then organizes, tabulates, depicts, and describes the data collection. It often uses visual aids such as graphs and charts to aid the reader in understanding the data distribution. Because the human mind cannot extract the full import of a large mass of raw data, descriptive statistics are very important in reducing the data to manageable form. When in-depth, narrative descriptions of small numbers of cases are involved, the research uses description as a tool to organize data into patterns that emerge during analysis [8–10]. Those patterns aid the mind in comprehending a qualitative study and its implications. In this article, we use random sample of employees and managers in the

institution of choice for Saudi Arabian Airlines in Jeddah, and numbered 23 employees and manager.

To achieve the objectives of the study were used to identify the development of a questionnaire study [2] to see the types and sources of information, the spectrum of relevance to the objectives of the study. The objectivity border was limited to the following objective limits how managerial decision-making process through sources and types of information resources used in managerial decision-making and the difficulties that they face it. On other hand, the process spatial border was confined to the border of Saudi Arabian Airlines in Jeddah Company.

5 NUMERICAL RESULTS

This study had used a questionnaire to gather data to provide managers with a score to identify their level of using the information resources. To ensure consistency the same tool was used for this research. Secondly, alongside qualitative differences between programs, a questionnaire offers the opportunity to look for statistically significant trends and differences. A questionnaire enables responses to be gathered from large numbers relatively quickly, and cost efficiently.

5.1 Demographic questions

The author use a questionnaire of [2] and distribute it to 23 manager given in Table 1 to be aware about the administrative position of the manager that he occupied in the organization and to which they belong

It was shown from Table 1 that 30.4% of the study sample branch managers, 21.7% of respondents said section managers and managing directors of 21.7%, 17.4% of respondents working in other positions. With regards to the number of the experience years, Table 2 shows the distribution of

Table 1. The distribution of the sample according to the administrative office.

Function	Repetition	Percentage
General Manager/President	2	8.7%
Deputy General Manager/Vice President	0	0%
Director of the Department	5	21.7%
Director of the Department	5	21.7%
branch manager	7	30.4%
Others	4	17.4%
Summation	23	100.0%

Table 2. The distribution of respondents by years of Experience in the field of management.

Experience period	Repetition	Percentage
Less than one year	3	13.0%
From one to five years	7	30.4%
From six to ten years	7	30.4%
From eleven to fifteen years	2	8.7%
From sixteen to twenty years	2	8.7%
More than twenty years	2	8.7%
Summation	23	100.0%

Table 3. The existence of library or information center.

Expression	Repetition	Percentage
Yes, there is information center	3	13%
No, there isn't information center	20	87%
Summation	23	100%

Table 4. Frequencies and percentages of the reality of the use of information Resources when making management decisions.

Expression	Always need		Often need		Sometimes need		Rarely need		Do not need at all	
	%	T	%	T	%	T	%	T	%	T
Identify the problem	78.3	18	0.0	0	13.0	3	4.3	1	4.3	1
Identify possible options to resolve the problem	65.2	15	13.0	3	8.7	2	13.0	3	0.0	0
Order possible options to resolve the problem and determine the optimal option	73.9	17	13.0	3	0.0	0	8.7	2	4.3	1
Apply the appropriate solution to the problem which was chosen	82.6	19	4.3	1	8.7	2	4.3	1	0.0	0
Evaluation of results	43.5	10	8.7	2	39.1	9	8.7	2	0.0	0

respondents by years of experience in the field of management.

It is shown from Table 2 that 30.4% of the study sample of their years of experience from one year to five years, 30.4% of respondents reported their years of experience from 6 to 10 years and 8.7% of respondents reported their years of experience from 11 to 15 members year, 8.7% of respondents reported their experience of 16 years to 20 years and 8.7% of respondents reported their years of experience more than 20 years. Regarding the important question about the existence of library or information center, it is shown in Table 3.

From Table 3, it is shown that 87% of the study sample there is no library or information center in the organization they work for 0.13% of respondents said no library or information center in the organization they work for.

5.2 Questions related to the use of information resources when making a decision

- The reality of the use of sources of information when making a decision:

From Table 4, it is shown the following important notes:-

- %78.3 of the study sample always need to identify the problem so that the administrative

decision is taken in each stage of the managerial decision-making.

- %65.2 of the study sample always need to identify possible options to resolve the problem is to take the administrative decision in each stage of the managerial decision-making.
- %73.9 of the study sample always need to arrange the possible options to resolve the problem and determine the best option is to take them to the administrative decision in each stage of the managerial decision-making.
- %82.9 of the study sample always need to apply the appropriate solution to the problem is to take the administrative decision in each stage of the managerial decision-making.
- %43.5 of the study sample always need to assess the results to be managerial decision-making in every stage of the managerial decision-making
- How to use information to make management decisions

Table 5 demonstrates how to use the administrative decision-makers of information

It is seen from the Table 5 that 60.9% of the study sample using the information in the daily decision-making routine to the same degree that they use in making strategic decisions.

Table 5. Various approaches of using the administrative decisions of the information.

Expression	Percentage	Repetition
The use of information in daily decision-making routine than using them in making strategic decisions	4.3%	1
Use the information in making strategic decisions than I use them in daily decision-making routine	34.8%	8
The use of information in daily decision-making routine equally used in making strategic decisions	60.9%	14
Summation	100%	23

6 CONCLUDED REMARKS

The role of information in decision making cannot be overemphasized. Effective decision making demands accurate, timely and relevant information. MIS provides accurate and timely information necessary to facilitate the decision-making process and enable the organizations planning, control, and operational functions to be carried out effectively. MIS also plays the crucial role of providing a wide range of streamlined options from which decision-makers are able to make their preferred choices and this ensures that whatever choices are made by decision makers, the outcome, more often than not, becomes positive.

Managers and business firms invest in information technology and systems because they provide real economic value to the business. The decision to build or maintain an information system assumes that the returns on this investment will be superior to other investments in buildings, machines, or other assets. These superior returns will be expressed as increases in productivity, as increases in revenues (which will increase the firm's stock market value), or perhaps as superior long term strategic positioning of the firm in certain markets (which produce superior revenues in the future). The results related by the use of sources of information when making a decision:

- The majority of the study sample always need (to identify the problem, identify possible options to resolve them and arrange them, and apply the appropriate solution) to the administrative decision is taken in each stage of the managerial decision-making.
- The majority of respondents said: they have accounted for 60.9% use the information in making routine daily decisions to the same degree that they use in making strategic decisions.
- The majority of the study sample always need (economic and financial information, and information on the market and competitors, and information on the application of informa-

tion technology in the work, and information on performance in the organization) to make decisions.

- The more sources of information that is used in an enterprise lines of decision-making (personal communication, consultation co-workers, specialized journals, and electronic databases, files institution), and less sources commonly used are (information services companies paid the price, and the library, whether outside the institution or within magazines).
- The more problems faced by members of the sample sometimes lie in access to information for decision-making (lack of a library or information center in the organization, and the lack of sufficient time to compile the information), and less problems they face (lack of information, and lack of enterprise files and documents to the organization).

REFERENCES

- [1] Ahmed Saleh Ahazzaamh (2009), "The role of information in decision-making system in government institutions: a field study in institutions for the general Irbid Governorate", Damascus magazine Economic and Legal Sciences
- [2] Jibril Arishi Abdulaziz Al-Omran (2003), "The reality of the use of information sources in the private sector in Saudi Arabia to support decision management representations: a survey in Riyadh".
- [3] Kamal Mustafa Roabh (2004), "Study the awareness of officials of Kuwaiti companies towards the use of strategic information: An Empirical Study", Arab Journal of Administrative Sciences, pp. 149-181.
- [4] Asefeh Asemi, Ali Safari, Adeleh Asemi Zavareh (2011), "The Role of Management Information System (MIS) and Decision Support System (DSS) for Manager's Decision Making Process", International Journal of Business and Management Vol. 6, No. 7; July 2011.
- [5] Mason, R.O. (1981)., Basic concepts for designing management information systems. In: Mason, R.O., & Swanson, E.B. (eds) Measurements for Management Decision, Philippines: Addison-Wesley.

- [6] Papows, J., (1998). Enterprise.com: Market Leadership in Information Age, London: Nicholas Brealey Publishing.
- [7] Power, D. J. (2002). Decision Support Systems: Concepts and Resources for Managers, Editor, DSSResources.COM. Quorum Books division, Greenwood Publishing.
- [8] G. Satyanarayana Reddy, Rallabandi Srinivasu, Spikanth Reddy Rikkula, Vuda Sreenivasa Rao, Management information system to help managers for providing decision making in an organization, International Journal of Reviews in Computing, 2009. IJRIC.
- [9] Lucey, T., 1997., Management Information Systems, London.
- [10] O'Brien, J. A. Management Information Systems: Managing Information Technology in the Inter-networked Enterprise, Boston: Irwin McGraw-Hill 1999.

Brain storming algorithm for coverage and connectivity problem in wireless sensor network

R.A. Ramadan

Computer Engineering Department, Cairo University, Giza, Egypt
On Leave at Hail University, Hail, Saudi Arabia

A.Y. Khedr

Systems and Computers Department, Alazhar University, Cairo, Egypt
On Leave at Hail University, Hail, Saudi Arabia

ABSTRACT: One of the famous problems in the field of Wireless Sensor Networks is the coverage. The coverage problem has many variations including the point, border, partial and full coverage to the monitored field. Different methodologies are used to solve this problem. Some of these methodologies are the Genetic Algorithms (Gas) and Circle Packing techniques. In this paper a new proposed algorithm is used that inspired from swarm intelligence entitled Brain Storming Optimization (BSO) algorithm. The algorithm is used with some of the normalization methods not only to solve the coverage problem but also the connectivity problem. BSO in this paper is treated as a multi-objective solution to the coverage and connectivity problems in WSNs. A set of case studies are simulated and compared to the GA. The results show that BSO is outperforming the Gas in terms of coverage and connectivity with small overhead that can be ignored.

1 INTRODUCTION

Coverage problem in WSNs takes many shapes including target coverage, area coverage, and barrier coverage. The idea behind the point coverage is to cover specific points or moving points in the monitoring field. The problem in its general form is NP-complete problem (Zhao & Gurusamy, 2008). Therefore, the authors of (Zhao & Gurusamy, 2008) tried to solve the problem using heuristic approach as maximum cover tree problem and show that it is an NP-complete problem. One more point coverage technique is proposed in (Gu et al., 2011) to monitor a moving object in the monitored field. They propose heuristic approximation algorithms. In Area coverage (Singh & Sharma, 2014), the main purpose is to cover the whole area of the monitored field with minimum number of sensors and prolonging the lifetime of the network as well. To do so, different scheduling algorithms are proposed including learning automata (Dietrich & Dressler, 2009) (Maggie & Xuan, 2011). The last class of coverage is the barrier coverage in which it could be defined as maximizing the detection of penetration of a border. This type of coverage is used mostly in applications like countries border monitoring.

Solutions to these problems also could be classified into centralized algorithms such as in (Cardei & Du, 2005) (Cardei & Wu, 2006) (Slijepcevic & Potkonjak, 2001) and distributed algorithms such as in (Tarik, Y. & Ezhan, K. 2010). In (Cardei & Du, 2005), the authors proposed an algorithm to monitor an object in the monitored field. The main idea behind their algorithm is to divide the sensor nodes into sets where each set can track/cover the target in the whole monitored area. This problem is a well-known problem under the title of maximum set cover problem. The authors in (Cardei & Wu, 2006) extended this work and proved that maximum set cover problem is NP-complete problem. In (Slijepcevic & Potkonjak, 2001), the authors took another approach where they divided the monitored area into fields that can be covered by the same set of nodes. The linear programming techniques were not far from targeting the coverage problem solution. For instance, in (Sung-Yeop & Dong-Ho, 2009), the authors proposed an Integer Linear Programming (ILP) for multiple target tracking with network lifetime extension. In (Tarik & Ezhan, 2010), a distributed was proposed for partial target coverage. In other words, they assume that the full coverage of the monitored field and/or the target is not required;

only the coverage is required with certain percentage. Sensors residual energy is utilized to check on the performance of the proposed algorithm.

There are many other variations to the coverage problem including coverage and connectivity (Mahmud & Fethi, 2015), k-coverage, connected k-coverage (Ramalakshmi & Radhakrishnan, 2015) and homogenous and heterogeneous coverage problems (Fatemeh & Ahmad, 2013). The target of all of these coverage problems is to prolong the lifetime of the WSNs. Some of the used methods are based on scheduling the sensors between sleep and wake up modes. Others try to benefit from the characteristics of the sensors such as initial energy and mobility.

The closest work to the work in this paper is the coverage problem stated in (Wengen et al., 2015). The problem is to maximize the coverage of certain area given a limited number of heterogeneous sensors. The heterogeneity of the sensors are in terms of their sensing ranges. The authors proposed special type of Genetic Algorithm (GA) with three different normalization methods which are Random, MinDist, and MaxDist. Random normalization is the same the original version of the GA where chromosomes are generated, crossed over, and mutated. In MinDist normalization, the chromosomes are rearranged to have the minimum distance between their corresponding genes then crossed over; on the other hand, in MaxDist, the chromosomes are rearranged to have the maximum distance between their corresponding genes then crossed over. Genes in these cases are the positions of the sensor nodes in the monitored field. The main considered objective function is the coverage. The problem with this approach is that the authors did not consider the energy of the sensors in their deployment model. In fact, the sensors energy could be an important issue since some of the sensor nodes might die in short period of time.

Throughout this paper, Brain Storming Optimization (BSO) will be used instead of GA taking into consideration another objective which is the network lifetime. BSO, is a new algorithm that appeared in 2011 and stated in (Wengen et al., 2015). The work in (Rabie, 2016) extended the original version of the BSO to include Fuzzy functions in the clustering phase of the algorithm and named the algorithm as Fuzzy Brain Storing Optimization (FBSO) algorithm. However, the used BSO in this paper is the modified version of the one proposed in (Wengen et al., 2015).

The paper is organized as follows: section 2 states the problem statement, section 3 explains the solution approach, the simulation experiments are presented in section 4 and finally the paper concludes in section 5.

2 PROBLEM STATEMENT

The coverage problem discussed in this paper belongs to the point coverage class of problems where sensors coverage model is considered as binary model. In the binary coverage model as given in equation (1), a point $p_j(x_j, y_j)$ will be fully covered if it falls within the sensing range r_i of a sensor $s_i(x_i, y_i)$.

$$F(x, y) = \begin{cases} 1 & d(s_i(x_i, y_i), p(x_j, y_j)) < r_i \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where $d(s_i(x_i, y_i), p(x_j, y_j))$ is the Euclidian distance between the two points $p_j(x_j, y_j)$ and $s_i(x_i, y_i)$, i is the sensor identification.

This model considers the sensor node sensing range as a disk/circle centered at the sensor itself. When the point $P_j(x_j, y_j)$ is covered the model value is equal 1 and 0 otherwise.

Then, the coverage problem could be defined as follows:

Given a set of heterogeneous sensors n where they differ in their sensing range r_p , communication ranges c_i and initial energy E_p . These sensors are supposed to be deployed in a monitored field A with length L and width W . the objective of the deployment is to maximize the coverage of the monitored field and prolonging the sensors lifetime by reducing the consumed energy in the network.

3 SOLUTION APPROACH

This section proposes to apply the Brain Storming Optimization (BSO) algorithm for the previously coverage problem. The algorithm follows the real brain storming strategy footsteps and the adapted version of the algorithm can be described as follows:

Step 1: Define the number of iterations (I_{\max}), number of ideas (D) to be initially generated, number of clusters (K_{\max}).

Step 2: Randomly generate D ideas.

Step 3: Evaluate the D generated ideas based on the coverage and consumed energy.

Step 4: Apply the Pareto dominance on the generated idea

Step 5: Apply K-Mean Clustering to cluster the generated ideas into number of clusters K_{\max} .

Step 6: Within each cluster, apply Crossover operation between the newly generated ideas and the old ones; select the best idea to replace the old one, if any.

Step 7: Rank the selected X solution from all clusters based on their coverage and lifetime and select the best no dominated ones (C_ideas).

Step 8: If the current number of iterations $> I_{max}$, go to Step 10.

Step 9: if the number of current ideas is less than D, randomly generate other (D- C_ideas) ideas and go to step 3.

Step 10: report the current idea and terminate.

The crossover methodology used in this paper inspired from (Junfeng, 2015) where three crossover methods are applied. The first crossover method is the random method where some of the sensors positions in each idea are exchange based on probability (Prob1). The second methodology is the MaxDist where sensors locations in each ideas are adjusted to have the furthest distance between the two corresponding sensors in each idea; then the random crossover operator is used. The third crossover method is based on the MinDist normalization where sensors positions are rearranged to have the distance between sensors in corresponding ideas are minimized; then the random crossover operator is applied.

K-means clustering is a famous clustering algorithm that tries to group objects based on their features into K classes where k is a positive number. The grouping is done by minimizing the distances between the data and the corresponding cluster centroid.

4 SIMULATION EXPERIMENTS

This section shows the performance of the BSO in solving the WSNs coverage problem compared to the Genetic Algorithm (GA) stated in (Wengen, 2015). However, the BSO tries to optimize the WSN not only based on the field coverage but also based on the network lifetime while GA in (Wengen, 2015) tries only to maximize the coverage. The network lifetime might got affected by sensors connectivity, used routing algorithm, and the number of messages to be sent through the network, and position of the sink node.

Monte Carlo sampling is used for coverage evaluation as given in (Wengen, 2015) considering X is the coverage area; therefore the coverage could be estimated as follows:

$$\bigcup_{i=1}^k \bigcup_{j=r_i}^{m_i-1} C_{ri}(x_j, y_j) \cap A \quad (2)$$

$$S(X) = \iint_A \lim_{L \rightarrow \infty} \frac{S(A)}{L} \sum_{i=1}^L I_x(x_i, y_i) \quad (3)$$

$I_x(\cdot)$ is defined as:

$$I_x(x, y) = \begin{cases} 1 & (x, y) \in X \\ 0 & (x, y) \notin X \end{cases} \quad (4)$$

In the literature different energy models can be found.

$$E_{Tx}(l \text{ bit} * dis) = \begin{cases} E_{elec} * l \text{ bit} + E_{amp} * l \text{ bit} * d^\alpha & \\ E_{elec} * l \text{ bit} + E_{fs} * l \text{ bit} * dis^2 & \text{if } dis^2 < d_0 \\ E_{elec} * l \text{ bit} + E_{mp} * l \text{ bit} * dis^4 & \text{if } dis \geq d_0 \end{cases} \quad (5)$$

We base our work on the first-order model described in (Ramalakshmi. & Radhakrishnan, 2015) where the transmitter or the receiver dissipates E_{elec} energy per bit to run the digital coding circuit, modulation circuit, and filtering of the signal circuit which are the radio electronic circuits before it is sent to the amplifier and dissipates E_{amp} in the power amplifier as shown in Figure 1. E_{amp} Varies according to the distance d between a transmitter and a receiver: $E_{amp} = \epsilon_{fs}$ assuming a free space model when $dis \geq d_0$ and $\alpha = 2$, while $E_{amp} = \epsilon_{mp}$ assuming a multi-path model when $dis \geq d_0$ and $\alpha = 4$, where $d_0 = \sqrt{(2 * \epsilon_{fs} / \epsilon_{mp})}$. Thus, to transmit an l bit-bit packet over distance dis, the radio expends:

and to receive this message, the radio expends:

$$E_{Rx}(l \text{ bit} * dis) = E_{elec} * l \text{ bit} \quad (6)$$

In these formulas, E_{Tx} is the transmission power, E_{Rx} is the receiver power. The radios are assumed to have power control and consume the minimal energy needed to reach the receiver.

The monitored area are assumed 800 mX800 m and the number of used sensors are assumed 100. The 100 sensors are assumed to be from three different categories in which they differ in their sensing ranges, communication ranges, and initial

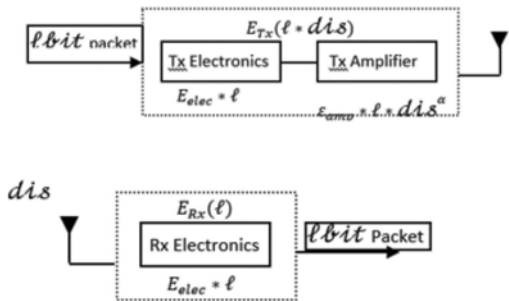


Figure 1. Radio energy dissipation model.

energy. The first category involves sensors with sensing range from 30 m, communication range of 60 m, and initial energy 0.5 j. The second category includes 50 m sensing range, 100 m communication range, and 1 j as initial energy. The sensing range of the last category is 60 m and the communication range is 90 m and the initial energy is assumed to be 1.5 j. The number of each category is selected based on a probability $P(\text{type})$ where each experiment runs for 10 times with different settings and different number of sensors per category; the average over the 10 runs are reported in the following case studies. A common energy model is assumed for all of them as can be seen in Table 1.

Case Study 1: Efficiency of BSO in terms of Convergence and Connectivity

This case study examines the performance of the proposed BSO and GA in 100 iterations. Figure 2 shows the coverage percentage of the BSO with different normalization methods (Random, MinDist, and MaxDist). It seems that BSO with MaxDist over performs the other two normalization methods in terms of coverage. However, the difference is not that much noticeable in small networks since it is almost 2%. In large WSNs, this difference becomes noticeable. At the same time, Figure 3 confirms the same results in which the GA MaxDist gives the highest coverage performance.

Another set of experiments are conducted to examine the comparison between the GAmix and BSOMax in terms. Out of Figure 4, it seems that the BSOMax is overperforming the GAmix by almost 15%. However, BSO still needs more enhancement since its coverage percentage is 75% of the

Table 1. Common energy model.

$E_{elec} = E_{DA}$	50 nJ/bit
E_{fs}	10 pJ/bit/m ²
E_{mp}	0.0013 pJ/bit/m ⁴

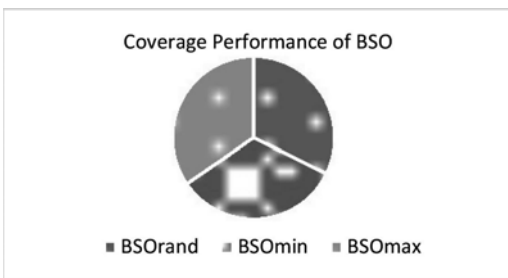


Figure 2. Coverage performance of BSO algorithms.

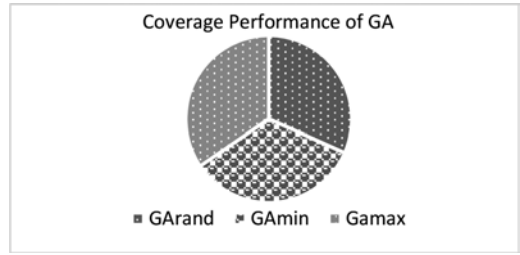


Figure 3. Coverage performance of GA algorithms.

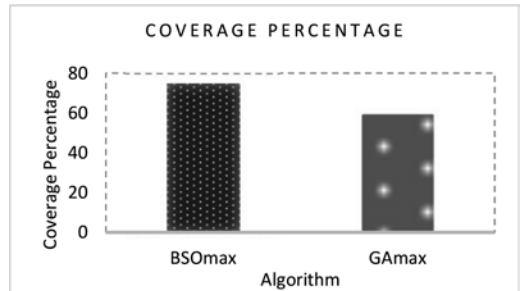


Figure 4. Coverage performance of the BSO and GA algorithms.

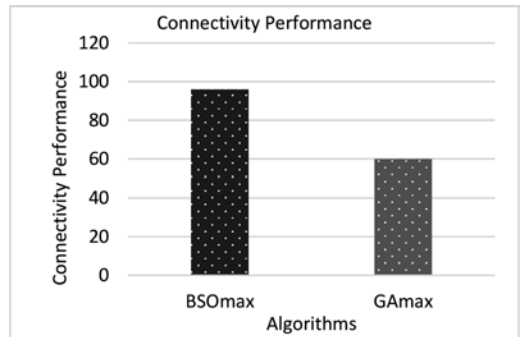


Figure 5. Connectivity performance of the BSO and GA algorithms.

monitored area. The results could be enhanced if the number of iterations are increased.

Figure 5 examines another issue that is not considered in (Wengen, 2015) which is the connectivity. The connectivity measure in this set of experiment is considered by how many nodes can be reached from any other node out of the total number of nodes. The figure shows that the BSO is much better than the GA by almost 20%. The reason behind that is the BSO is considering the connectivity as one of its objectives.

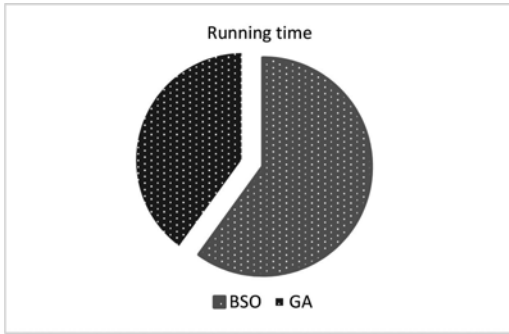


Figure 6. Running time for the BSO and GA algorithms.

Case Study 2: Running Time Evaluation

Looking closely at the proposed BSO algorithm and GA, it is found that BSO needs only two things more than the GA which are the clustering and dominating points check process. K-mean Clustering is used here in the BSO which costs $O(Kt_{\text{dist}})$ where K is the number of clusters and t_{dist} is the number of objects and t_{dist} is the time to compute the centroid. Therefore, the BSO as given in Figure 6 is taking more running time than the GA. However, this amount of increase in the running time is not that much and it is within the linear complexity.

5 CONCLUSION

The research in this paper targeted solving the coverage problem in wireless sensor networks. The proposed algorithm is known as the Brain Storming Optimization (BSO) is utilized for the coverage problem. Different variations are considered, especially in case of normalizing the ideas. BSO solution is compared to the traditional GA and the results show that it outperforms the GA in terms of coverage and connectivity.

REFERENCES

- Cardei, M. & D-Z.Du. 2005. Improving Wireless Sensor Network Lifetime through Power Aware Organization. in *ACM Wireless Networks*, 11: 333–340.
- Cardei, M., & Wu, J. 2006. Energy-Efficient Coverage Problems in Wireless Ad Hoc Sensor Networks. *Computer Communications*, 29(4): 413–420.
- Dietrich, I. & Dressler, F. 2009. On the life-time of wireless sensor networks. *ACM Trans. Sen. Network.* 5(1).
- Fatemeh, M. & Ahmad, K. 2013. Coverage Problem in heterogeneous Wireless Sensor networks. *European Scientific Journal* September 2013 edition, 9(27). ISSN: 1857–7881 (Print) e - ISSN 1857–7431.
- Gu, Y. Zhao, BH., and Ji YS. 2011. Theoretical treatment of target coverage in wireless sensor networks. *Journal of Computer Science and Technology*, 26(1): 117–129.
- Junfeng, C. Shi, C. Yang, C., Yingjuan, X. Yuhui, S. 2015. Enhanced Brain Storm Optimization Algorithm for Wireless Sensor Networks Deployment, *Advances in Swarm and Computational Intelligence*, series Lecture Notes in Computer Science 9140: 373–381.
- Maggie, X. & Xuan, G. 2011. Maximum lifetime coverage preserving scheduling algorithms in sensor networks. *Journal of Global Optimal*, DOI 10.1007/s10898-010-9636-3.
- Mahmud, M. & Fethi J. 2015. An Iterative Solution for the Coverage and Connectivity Problem in Wireless Sensor Network. *The 6th International Conference on Emerging Ubiquitous Systems and Pervasive Networks*.
- Rabie A, “Fuzzy Brain Storming Optimization (FBSO) algorithm” accepted at *International Journal of Intelligent Engineering Informatics*, 2016.
- Ramalakshmi, R. & Radhakrishnan, S. 2015. Connected k-Coverage Topology Control for Area Monitoring in Wireless Sensor Networks. *Wireless Personal Communications*, 84(2): 1051–1067.
- Singh, A. & Sharma, T. 2014. A survey on area coverage in wireless sensor networks. *International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, 829–836.
- Slijepcevic, S. & Potkonjak, M. 2001. Power Efficient Organization of Wireless Sensor Networks. At the *ICC, Helsinki, Finland*, 2: 472–476.
- Sung-Yeop, P. & Dong-Ho C. 2009. Power-Saving Scheduling for Multiple-Target Coverage in Wireless Sensor Networks. *IEEE COMMUNICATIONS LETTERS*, 13(2): 130–132.
- Tarik, Y. Ezhan, K. 2010. A distributed activity scheduling algorithm for wireless sensor networks with partial coverage. *Wireless Networks*, 16: 213–225, DOI 10.1007/s11276-008-0125-2.
- Wengen, G. Qigong C. Ming J. Yunfei, L. and Shinong, W. 2015. The Optimization of Genetic Algorithm in Wireless Sensor Network Coverage. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 8(1): 255–264.
- Zhao, Q. & Gurusamy, M. 2008. Lifetime maximization for connected target coverage in wireless sensor networks. *IEEE/ACM Trans. Netw.* 16(6): 1378–1391.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Evaluating IPTV network performance using OPNET

Eman S. Sabry

*Department of Electronics and Electrical Communication, Higher Institute of Engineering,
El Shorouk Academy, Cairo, Egypt*

Rabie A. Ramadan

*Cairo University, Cairo, Egypt
Hail University, Hail, Saudi Arabia*

M.H. Abd El-Azeem & Hussien ElGouz

*Arab Academy for Science & Technology and Maritime Transport-College of Engineering & Technology
Electronics & Communication Engineering Department, Cairo, Egypt*

ABSTRACT: Internet Protocol Television IPTV is a new television platform launched to overcome the deficiencies of WebTV. IPTV is considered a perfect platform for new ideas and concepts. This paper provides a deep insight into IPTV technology with existing protocols, evaluating the impact of uncompressed video delivery at different data rates on network performance. The second contribution of this paper is to evaluate IPTV network performance concerning different application issues related to the delivery of coded Single View (SV) and Multiple Views (MV) videos in the currently used H.264 codecs format as a part of Optimized Network Engineering Tools (OPNET) simulator. Moreover, the preamble paper includes an IPTV network performance evaluation subjected to the coded video channels delivery in the recently introduced High Efficiency Video Coding (HEVC) standard format at the same resolution and over the same network. Thus, it could be easy to find out the most instrumental codec for IPTV revolution in order to satisfy both end users and service providers satisfaction.

1 INTRODUCTION

Generally, Digital Generation represents the state of the art of modern television technology, defeating analogue TV with its quality. Moreover, Internet TV entertainment achieves a realized progress that's why the most common entertainment Internet TV is IPTV. IPTV is a very stringent QoS and QoE technology since video contents are delivered over dedicated private and secured fixed geographical area networks (ITU—IPTV FG 2008). Lossy compression can be used for audio/video AV in which it approximates the media data rate dramatically. H.264/AVC technology is considered the most popular and dominant lossy compression method used by commercial IPTV providers. Typically, MV contains two views of a video taken from different perspectives, whereby each view consists of a sequence of video frames (pictures). These two different views could be displayed to give viewers the perception of depth. This commonly refers to Three-Dimensional (3D) video; to provide 3D video services over transport networks, it is required to have efficient video compression (coding) techniques and sufficient transport mecha-

nisms to accommodate the large video data volume from the two views on limited bandwidth transmission links. MV is an amendment to the H.264 (MPEG-4 AVC) video compression standard.

Hence, efficient coding techniques for MV video have been extensively researched in recent years but it is still an open area for research. Therefore, the industry keeps looking for the global benchmark for video compression which is ITU and its partners since ITU-T H.264 has underpinned expansion and rapid progress. HEVC or ITU-T H.265 was jointly developed to double the video data compression ratio as compared to its predecessor ITU-T H.264 / MPEG-4 Part 10 - Advanced Video Coding—(AVC) at the same level of video quality even better. HEVC opens the future door for video transmission only using half of the bandwidth (bit rate) compared to its predecessor, which currently accounts over 80 percent of all web video. H.265/HEVC, the further advance video coding is emerged as video coding standard and 3D video coding serving multimedia communications. The rest of the paper is organized as follows: Section (2) contains the Literature work explaining recent relevant IPTV research papers. Section (3) shows

required IPTV nodes and main architecture elements for implementation of IPTV network. Section (4) elaborates different networking statistics that will be included. Section (5) shows scenarios and results. Section (6) draws the conclusion. Section illustrates helpful references.

2 LITERATURE REVIEW

Recently, several works have been introduced based on performance studies of video streaming over IPTV networks. For example, in (Moughit & Badri, 2013), the authors study the effect of Cisco Group Management Protocol (CGMP) on the IPTV performance in terms of throughput and delay. However, the authors ignored the details of the IPTV architecture including routing and queuing mechanisms; in addition, the paper works on a very simplified version of IPTV to utilize only two scenarios with CGMP enabled and disabled in terms of delay and throughput only. In (Singh & Amit, 2013) authors result from IPTV VoD transportation assessment over WIMAX that Packet Delay Variation (PDV), packet End-to-End (ETE) delay and delay decreases with little load increase by the increase of WIMAX mobile subscriber mobility. Similarly, in (Hamodi & Thool, 2013) authors introduce a study for IPTV performance analysis over WIMAX broadband access technology. Indeed previous valuable effort is communicating IPTV assessment from the perspective of subscriber mobility only. System evaluation judgment lacks system examination subjected to variety of practical application effects to assimilate more channels and users in both objective and subjective merits.

In (Maraj, Shehu & Mitrushi, 2011) authors try to assess QoS parameters and QoS requirements for delivering IPTV services, as transmitting different QoS sensitive services. They want to design some controlling mechanisms for solving different problems that might occur in network in case of delays, losses, etc. using Fuzzy Logic Controller (FLC). Indeed, this assessment and enhancement is based upon assumption as it lacks the practical network deployment.

In (Kokoška, Handriková, & Valiska, 2014), the authors describe a couple of available network simulators and differentiating between them in order to decide on the proper one for better realization to the whole IPTV QoS process. Unfortunately, the survey was a shallow in terms of not taking into consideration simulators details and network settings. In addition, the survey did not go deeply into the implementation of IPTV on such simulators. However, our paper explains different OPNET capabilities that might help in implementing IPTV

such as the available queuing system library, measuring delay and jitter, estimating the throughput, and capability of modifying network elements. In addition, it implements the IPTV features for the purpose of assessment. Our contributions in this paper are IPTV network performance evaluation subjected to uncompressed multi-casted video at different data rate, and subjected to coded videos either in H.264 or H.265 formats as to find out the most proper codec format helps in the revolution of the introduced technology.

3 IPTV NETWORK CONFIGURATION OVER OPNET

In this section, different components of IPTV that will be implemented using OPNET will be detailed.

3.1 IPTV hardware

Any IPTV system is made up of four major (elements) domains, all are generic and common to any vendor's (or combination of vendors') infrastructure as shown in Figure 1. The four network domains involve IPTV Data Centre or Video Head end that is responsible for capturing or acquiring video from different sources involving receiving, decoding and decrypting multimedia contents, core network that is located at the network edge connected to the access network to transport the encoded group of channel forwarded from the video head end, access network (IPTVCDs) that is responsible for decoding and processing the incoming IP based video stream allows users to access IPTV services including DSLAM (DSL Modem), and finally Home network that is responsible for IPTV service distribution throughout the home; many different types of home networks exist including TV sets and STBs. However, IPTV requires a very robust high bandwidth home network that can only be accomplished today using wire line technology.

The implemented network modeling IPTV is shown in Figure 2 using OPNET including the following nodes: IPTV_Headend_Video node,

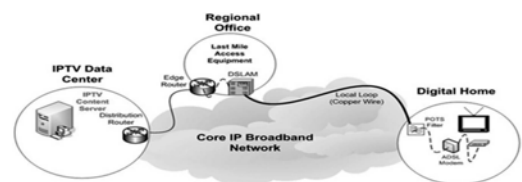


Figure 1. IPTV main components.

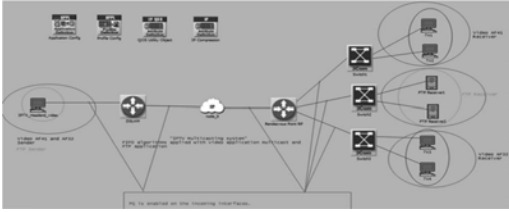


Figure 2. OPNET IPTV model.

DSLAM node, and a Rendezvous Point (RP) multicasting node. The rest of architecture nodes in Figure 2 are switches, routers, and TV sets. Be notice that, a RP is a core router at which multicast domain packets from the upstream source and join messages from the downstream routers are “rendezvous”. It is important to know that other network’s routers do not need to know the source address of every multicast group after RP configuration. Therefore for correcting operation, the rest of the network routers supporting multicast have to know RP address and this is performed through IP addressing of any active RP interface. In addition, QoS Utility Object node is used to assign quality of service schemes and queuing algorithms, and IP compression node to perform data gram compression.

3.2 IPTV networks protocols

The following are the protocols for IPTV networks:

- Routing protocols: OPNET offers variety of information routing protocols that could be easily configured whether the network is managed by the same Autonomous System (AS) (intra-AS routing protocols) or between ASs (inter-AS routing protocols). Routing Information Protocol (RIP) and/or Open Shortest Path First (OSPF) could be assigned.
- IP Multicasting: OPNET modeler supports IP multicast including Internet group management protocol IGMP and Protocol Independent Multicast-Sparse Mode (PIM-SM). IGMP is used by hosts and adjacent routers to establish multicast group memberships. A TV set transmits IGMP-join/leave messages to notify the upstream equipment by LEAVE-ing one group and JOIN-ing another channel. PIM-SIM is multicasting routing protocol that explicitly builds shared trees rooted at a RP per group and optionally creates shortest-path trees per source.
- IP Compression: Compression decreases the packet size by compressing certain portions of the datagram; this differs to the well-known

video compression. The choice of datagram compression type depends upon network size.

- QoS (Quality of Service): As IPTV subscribers expect a specific viewing quality level, IPTV service provider uses differentiated services (DiffServ) protocol for specifying and controlling network resources and bandwidth for packet transportation. Using OPNET to enable examination of IPTV offered QoS support, by traffic marking specifying traffic with different classes that are served according to queuing priority. Priority Queuing (PQ) and Custom Queuing (CQ) schemes are deployed to queue traffic of different classes (Sethi, & Hnatyshin, 2013).

4 IPTV NETWORK PERFORMANCE CHARACTERIZATION

This section assesses different performance metrics key issues related to video transmission over IPTV. Regarding the Internet services, QoS measurements parameters usually are as follows:

- Packet Delay Variation (PDV): shows variance among ETE that video packets experience to. OPNET permits collection of Global PDV statistic within created network that is recording data from all network’s nodes.
- Packet End-to-End (ETE) delay (sec): average time counted to send a video application packet to a destination node application layer, it could be computed using equation (1).

$$D_{ETE} = Q(d_{proc} + d_{queu} + d_{trans} + d_{prop}) \quad (1)$$

Where Q indicates network nodes elements between IPTV Headend and TV set, d_{proc} is the network node processing delay, d_{queu} is the network node queuing delay, d_{trans} is the packet transmission time between two network elements on a communication link, and d_{prop} is the propagation delay within network link.

According to equation (1), counted ETE delay is seriously affected by network nodes and links influencing both QoS and user’s QoE. Hence, according to several fulfilments in this tenor ETE delay range has been identified. In which if delay is larger than 1 second; it produces bad QoS toward unacceptable service from end user’s QoE, while for one way communication if the delay is less than 200 ms better QoS is produced and hence acceptable service from end user’s QoE side (ITU—IPTV FG 2008).

- Packet jitter (sec): Measuring the difference between End-to-End delays of two consecutive packets, OPNET counts absolute value of the

recorded difference. For satisfying user's QoE the jitter delay for one way must be less than 60 ms on average and less than 10 ms in ideal case (ITU—IPTV FG 2008).

5 SIMULATION RESULTS

In this section different experiments are conducted for the purpose of measuring the IPTV performance using OPNET over IPTV modeled network in Figure 2. The following subsection explains how the background traffic that will be used in the later experiments is generated. The following sections are the performance experiments that conducted in this paper.

5.1 Video coding traffic generation (Background traffic generation)

OPNET provides conversation pair traffic as a way to traffic modeling over network that could be imported from outside sources. It is injected at different network layers as IP traffic flows according to a layer where data is deployed. This section suggests the way for coded video ingestion over network using video traces as text traffic file in .tr1 OPNET format to be imported as IP traffic flow. Using the huge coded video library traces in (Trace Files, 2015) for different coding standards, each collected trace file includes number of coded frames within each Group of Pictures (GOP) and their corresponding bytes/sec involved in these frames. These traces are converted into .tr1 text file format that is constructed of several traffic lines; each traffic line defines five different header fields. These header fields per each line define source node, destination node, time window (sec), number of transmitted frames within defined time window, and the corresponding bytes/sec included within these frames.

In order to evaluate network performance subjected to coded SV and MV in H.264 codec format, traces of the same movie sequence are collected and converted into the previously mentioned .tr1 text file format to be imported through network as IP traffic flow modeling two coded TV channels. Moreover, as to examine and to evaluate network performance subjected to coded video in the new codec HEVC format, H.265 coded video trace is also collected and converted into .tr1 similarly as its corresponding to H.264 format, as to model other coded TV channel in HEVC. Thus, three coded TV channels will be imported separately over IPTV network as an IP traffic flow in .tr1 OPNET format. The three collected traces are for the same movie sequence, however SV and MV coded videos are of resolution 1920×1088 and of FPS 48, while HEVC coded video is of resolution 1920×1088 and 30 FPS.

5.2 Experiment 1: Evaluating network IPTV with uncompressed video transmission at different data rates

Through Figure 2 network, two video applications are defined through Application Config node named Video Conferencing (AF41) and Video Conferencing (AF32). The two mentioned video applications are distinguished with Differentiated Services Code Point (DSCP). Video Conferencing (AF41) has the following characteristics: 1) Video packets have lower drop precedence than of that in Video Conferencing (AF32) application and marked to take higher priority through deployed QoS queuing priority schemes, 2) Video Conferencing (AF41) models a TV channel with frame size information 352×240 that is transmitted with frame inter-arrival time of 30 Frame/Second (FPS) and Type of Service (ToS) of AF41. On the other hand, Video Conferencing (AF32) models another TV channel with frame size information 128×240 that is transmitted with frame inter-arrival time of 15 FPS and ToS of AF32. Two corresponding user profiles are defined through Profile Config node named Video_AF41, and Video_AF32. According to attributed profile Config node all defined users profiles are configured in simultaneous operation mode as to be delivered together. Moreover, the assigned start time for each defined user profile will start to be collected after simulation start time by approximately 1 min 75 s.

Video_AF41 and Video_AF32 are attributed to the IPTV_Headend_Video node only to represent IPTV source node. The two defined applications are multi-casted over IPTV network with two corresponding multicasting group addresses which are: 224.233.24.231 and 224.233.24.232. In which 224.233.24.231 multicast address is associated with Video_AF41 service and 224.233.24.232 multicast address is associated to Video_AF32 service.

Four receiving nodes are configured in IPTV network as shown in Figure 2 as follows: 1) Two TV sets (TV1 and TV2) sharing the same multicasting address group for Video Conferencing (AF41) delivery, and 2) Two TV sets (TV3 and TV4) sharing the same multicasting address group for Video Conferencing (AF32) delivery. In addition, PIM-SM protocol is enabled in network routers (DSLAM and RP) for saving IPTV bandwidth save.

In order to examine IPTV network performance with the delivery of multi-casted video channels with high frame rates and low frame rates. Then in other separate DES simulation phase over Figure 2 network applications are settled to have Video Conferencing (AF41) frame inter-arrival time is reduced to 15FPS, and the Video Conferencing (AF32) frame inter-arrival time is reduced to 10 FPS.

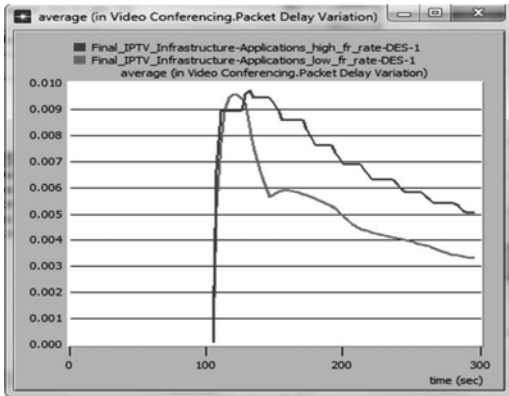


Figure 3. Global PDV associated with the delivery of video channels at high and low frame rates.

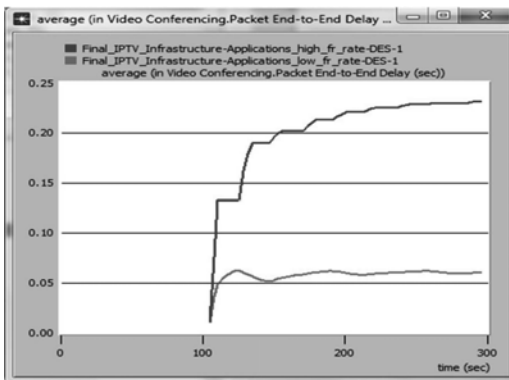


Figure 4. Global packet ETE delay associated with the delivery of video channels at high and low frame rates.

Figure 3 shows the global PDV with time in both high and low data rates. As given in the Figure 3, it seems that PDV for the high data rate is higher than in the low rate by almost 25%. However, at minute 2, the PDV of the low rate shows a peak that is quickly disappeared later. This is due to the network handling to the generated packets in terms of buffering and processing at low rate.

Figure 4 illustrates global ETE delay; illustrating the corresponding video packets carrying high frame rate channel experience higher ETE than of that with low frame rate. For instance, at 5 m 0 s a time video packet of high data rate achieves increase almost about 180 ms in ETE delay over their corresponding of low data rate.

So, to conclude the results from the previous Figures, it is obvious from ETE delay that users with high data rate might experience bad quality in data transfer. However, the results show that

it is almost three times more delay than the one with low data rate. At the same time, what makes it worth is that due to the high PDV values in high data rates packets may arrive out of order as well. Entirely this reflects the importance of video compression; not only its contribution in nodes buffering reduction since limited memory and speed at any receiving node but also its impact on quality of received TV video channels as concluded from previous experiment.

5.3 Experiment 2: Evaluating network performance in case of coded video transmission

The main objective behind this experiment is to evaluate IPTV network performance subjected to the delivery of SV and MV video sequence coded in H.264 format, and to the delivery of coded video in the new introduced H.265 codec format, and modeling three coded TV channels in two different codecs standards over IPTV network. This experiment uses the same previously created network setup in sub-section (5.2) with the following modifications:

- Failing all receiving network nodes in Figure 2 except TV1, as it is assumed exist for only a receiving node over network.
- Admission of previously prepared SV H.264 traffic file (SV H.264 .tr1 text file) in sub-subsection (5.1) subrogating uncompressed video channels, to model the delivery of only one SV TV channel in H.264 format over the same network.

It should be noted that any imported traffic file will be matched with network source and destination nodes by the OPNET importing machine. OPNET traffic center ensures this importation indicating the existence of only one IP traffic flow from IPTV source node to the only receiving set (TV1) receiving SV H.264 coded channel. Then, in two separate DES simulation phases the other coded TV channels are imported as the following:

- Firstly, admission of previously prepared MV H.264 traffic file (MV H.264 .tr1 text file) in sub-subsection (5.1) subrogating SV H.264 video channel, to model the delivery of only one MV TV channel in H.264 format over the same network.
- Secondly, admission of previously prepared H.265 traffic file (H.265 .tr1 text file) in sub-subsection (5.1) subrogating MV H.264 video channels, to model the delivery of only one coded TV channel in H.265 format over the same network.

Moreover, the background traffic delay (sec) is set to 150 sec defining when results will start to be

collected. Hence, all coded traffic channels admission will be started over network nodes after simulation start time by approximately 2 min.

Figure 5 illustrates average received traffic (bps) in case of the delivery of coded SV and MV channels in H.264 format. Figure 5 shows, in case of the delivery either SV or MV coded channels, a peak at 4 m 35 s that is achieved loading network of maximum traffic received of about 550 Kbps and 750 Kbps respectively, and then goes down again. This entirely reflects a number of variations for transmitted bytes/sec included within transmitted frames in these imported traffic files. Moreover, the figure shows that received traffic (bps) for coded MV channels have the same external modality as their corresponding in case of coded SV; this could be induced as both channels are for the same movie sequence. However, they differ in data rate as MV channel loads network with maximum data rate (bps) of about 1.5 times over SV H.264 inspected in terms of traffic received.

Figure 6 illustrates the average traffic received (bps) for the delivery of single HEVC video channel over the same network, it involves the collected received traffic that has the same external modality as their corresponding in case of coded SV shown in Fig. 3 (a). This could be induced as they are used

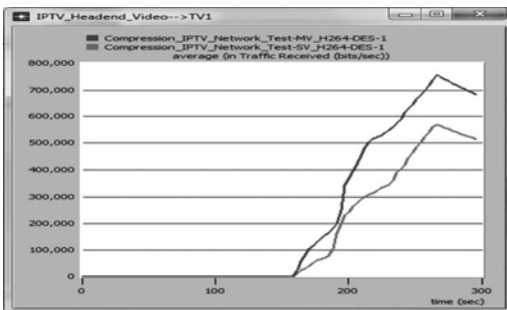


Figure 5. Traffic received (bps) with the delivery of coded SV and MV channels in H.264.

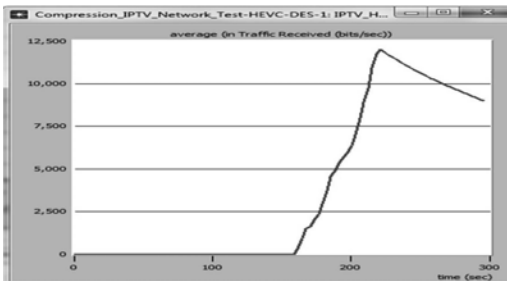


Figure 6. Traffic received (bps) with the delivery of coded channel in H.265.

for the same movie sequence; however they differ in data rate as H.265 coded channel achieves reduction in data rate (bps) over that for coded SV and MV channel inspected in terms of traffic received.

Figure 7 illustrates corresponding PDV with time in case of the delivery of coded SV and MV channels in H.264 format, showing that the PDV increase is roughly linear with time in case of SV coded channel. On the other hand in case of MV delivery, PDV is increased curvy with time unlike of that in case of SV delivery. This is due to network capability to handle the generated packets in terms of buffering and processing, reflecting the impact of high data rate involved in MV channel on network performance. Hence, MV packets deliveries are more vulnerable in the arrival of different ETE delays and in out of order as compared with their corresponding of SV TV channels.

Figure 8 illustrates PDV with time in case of the delivery of HEVC TV channel, it shows that with the delivery of HEVC coded channel PDV is roughly increasing with low slope as compared to that for H.264 SV shown in Figure 7.

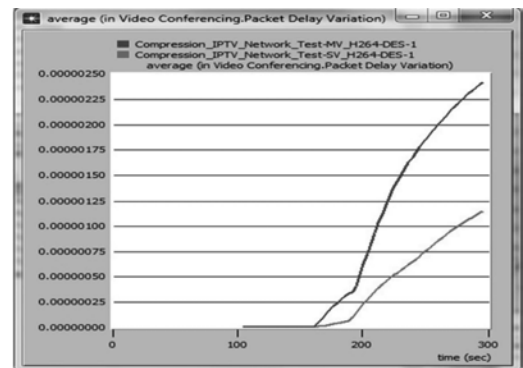


Figure 7. PDV with the delivery of coded SV and MV channels in H.264.

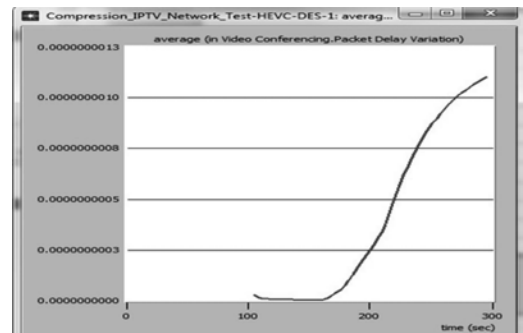


Figure 8. PDV with the delivery of coded channel in H.265.

As there is correlation between subjective and objective merits, it could be easy to estimate the quality of received video according to resulted performance networking terms. The most important parameters warranting the estimation of received video quality is ETE delay and packet jitter. Thus, Figure 9 shows packet jitter (sec) resulted from the delivery of coded SV and MV TV channels. The figure shows that over majority of time MV packets form higher picks in packet jitter than SV in which coded MV channel achieves maximum packet jitter of 38 ms; while coded SV channel achieves 21 ms maximum packet jitter. The resultant comparison between each case involves the delivery of MV in H.264 format that achieves increase of packet jitter by approximately 55 ms over of that for coded SV channel over the same network.

Figure 10 illustrates that maximum achieved Packet jitter for H.265 delivery is about 22 ms. Showing great reduction in packet jitter with their corresponding in case of H.264 coded channels.

Figure 11 shows packet ETE delays for the delivery of SV and MV coded TV channels and coded video channel in HEVC format. The figure shows that all cases have high external modality, as they are for the same movie sequence. However, coded

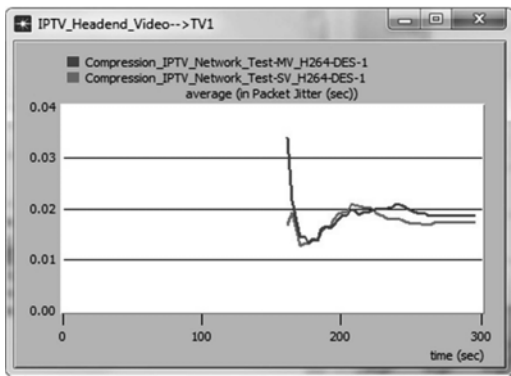


Figure 9. Average packets jitter in case of SV and MV H.264 video channels.

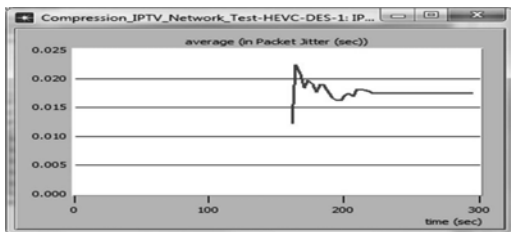


Figure 10. Average packets jitter in case of coded H.265 channel.

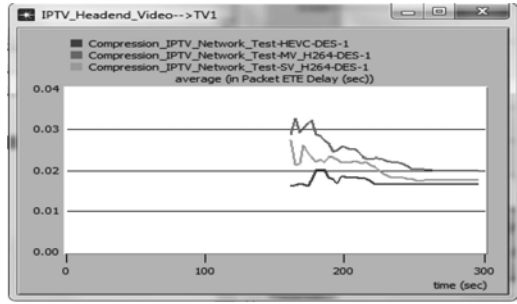


Figure 11. Average ETE delay included with the delivery of H.265 video, SV H.264 video, and MV H.264 video.

MV channel achieves maximum ETE delay of 34 ms coded SV channel achieves maximum ETE delay of 25 ms; while HEVC achieves maximum ETE delay about 20 ms. With the comparison of network performance subjected to the delivery of SV and MV coded channels; it is found that MV coded channel achieves increase in ETE delay by approximately 75 ms with their corresponding for SV coded channel over the same network. So, conclusion raised highlights video compression impact on quality of received service with respect to of that within experiment 1 simulation results. However, H.264 video standard has recorded several limitations on the increase of the number of transmitted HD or MV video channels to satisfy user and service providers recommended objectives. Moreover, evaluating of network performance in case of HEVC in terms of ETE delay and packet jitter, it is found that users receiving TV channels in HEVC format will expect very high video quality as compared to other two cases, comprising the great enhancement achieved by HEVC on network performance.

6 CONCLUSION

This paper introduces a detailed study for evaluating IPTV network performance subjected to multi-casted uncompressed video with assigned standard protocols behaviour at different data rates, highlighting the impact of video compression on QoS. This paper also provides evaluation for IPTV network performance subjected to the delivery of coded SV and MV in H.264 format and coded video in the new developed HEVC. From the results explained in the paper the following remarks could be concluded:

- Coded channel in H.265 format achieves reduction in maximum traffic received (bps) as compared to SV and MV in H.264 format of about

46 times and 63 times respectively. Moreover, Coded channel in H.265 format achieved reduction in maximum PDV as compared to SV in H.264 format of about 92300%.

- Coded SV channel in H.264 format achieves an increase in total ETE delay and packet jitter for all packets of about 300 ms and 260 ms respectively over their corresponding in HEVC format. Meanwhile coded MV video channel in H.264 achieves an increase in total ETE delay and packet jitter for all packets of about 410 ms and 300 ms respectively over their corresponding in HEVC. Moreover, HEVC achieves reduction in maximum ETE delay over MV H.264 of about 16.5 ms, while achieving reduction of about 14 ms over SV H.264 at the same time instance.

Thus, evaluation shows the great enhancement achieved by the behalf of the new codec on IPTV system as to satisfy user's recommendations in terms of both QoE and QoS.

REFERENCES

- Hamodi, Jamil M & Thool, Ravindra C, January 2013, Performance Evaluation of IPTV over WiMAX Networks Under Different Terrain Environments, International Journal of Engineering Inventions, Volume 2, pp. 21–25.
- ITU—IPTV Focus Group Proceeding, 2008, International Telecommunication Union.
- Kokoška, Rastislav, Handriková, Janka, and Valiska, Ján, 2014, Software Network Simulators for IPTV Quality of Services, Acta Electrotechnica et Informatica, Vol. 14, No. 1.
- Maraj, Arianit & Shehu, Adrian & Mitrush, Rozeta Miho, 2011, Studying of different parameters that affect QoS in IPTV systems, Proceedings of the 9th WSEAS International Conference on Telecommunications and Informatics.
- Moughit, Mohamed & Abdelmajid, Badri, February 2013, A multicast IPTV bandwidth saving bandwidth, International Journal of Computer Applications (0975-8887) Volume 64, No. 14.
- Sethi, Adarshpal S, and Hnatyshin, Vasil Y, 2013, the Practical OPNET® User Guide for Computer Network Simulation, New York.
- Singh, Gurmeet & Amit, Grover, 16 July, 2013, Simulation and analysis: The effect of mobility on IPTV (VOD) over WiMAX using OPNET, International Journal of Physical Science, Vol. 8 (26), pp. 1401–1407.
- Trace, 2015, Files [Online]. Available: <http://trace.eas.asu.edu/tracemain.html>.

Time Based Weighted Shortest Path Movement Model (TBW-SPMM)

Ahmed B. Altamimi

*Computer Engineering Department, College of Computer Science and Engineering,
University of Hail, Hail, Saudi Arabia*

ABSTRACT: Random movement models do not capture a realistic nodes movement as in real life scenario. The importance of a realistic movement models is to accurately predict networks performance, including the Internet of Things (IoTs) environment where more than one network are interacting. This paper proposes a movement model that can characterize the nodes movement in IoTs sub-networks, including intermittently connected network and wireless sensor network. The realistic characteristic in the proposed model is compared against real-world experiments traces for validation.

1 INTRODUCTION

Internet of things is defined as a network that allows sub networks including sensor, mobile, and vehicular networks to communicate with each other. The integration between different networks in IoT can bring the benefits and the characteristics of each network to the IoT environment. For example, sensor networks are mainly used for sensing and collecting data. Additionally, it is usually considered as static networks. On the other hand, mobile or vehicular networks are often employed for unicast communication between nodes or vehicles. Additionally, it is known for its mobiles behaviors. Bringing these different characteristics and benefits in one environment is attractive. The attractiveness of IoTs environment is dependable on the performance of its sub networks. The performance of a network cannot be accurately measured without accurate nodes movement modeling. This brings the importance of the discussed model in this paper.

The movement models are classified into random based and map based models. Nodes move on random speeds and directions in random based models. An example of such model is random waypoint model (Johnson et al. 1996). In map based model, nodes movements are constrained within predefined routes on the map. Shortest Path Movement Model (SPMM) (Keranen et al. 2007) is an example of map based model. Recently, many research activities have focus on map based model (Shahzamal et al. 2014) since it is considered to be more realistic than random models (Shahzamal et al. 2014).

This paper proposes a map based movement model, namely, Time Based Weighted Shortest Path Movement Model (TBW-SPMM). It considers the period or the time of nodes movement and the weight of where a node is headed to, the weight is determined by how busy the next node location is, as two factors to determine the nodes movement. This ensures a realistic movement modeling as it is validated against a real-world trace experiment.

The paper is divided to five sections. Next section presents related works. The TBW-SPMM is discussed in the third section. The simulation result is presented in section 4. Section 5 concludes the paper.

2 RELATED WORK

Movement models are classified to random (Shahzamal et al. 2014) and map (Shahzamal et al. 2014) based. Examples of random based models: a) Random Walk (RW) (Walker et al. 2008) where nodes move on random speed and direction, b) Random Waypoint Model (RWP) (Petz et al. 2009), where a node stay for a pause time before it heads to a new destination with random speed and directions, c) Random Direction (RD) (Zaninetti et al. 2008), a node reach the boundary of the simulation area before it heads to the new destination with random speed and directions, as well. Random based model shows unrealistic behavior of nodes movements as reported in (Keranen et al. 2007). This leads to the development of map based to achieve more realism behavior of nodes movements.

Map based models examples are: a) Map-Based Mobility Model (MBM) (Keränen et al. 2010), where nodes move on random speed and direction on predefined map routes, b) Shortest Path Movement Model (SPMM) (Keränen et al. 2007), where nodes behave similarly as in MBM but with shortest path between two points is taking in consideration. SPMM assumes the existing of Points of Interests (PoIs) in the map, such as restaurants or shopping mall, as destinations for moving nodes. The proposed movement model weight the PoIs from being the busiest PoI to the least busy one based on the number of visitors in each one. The busy PoI attracts more nodes than less busy PoI location. Finally, nodes move to a particular PoI also based on the time of the year/week/day.

3 TIME BASED WEIGHTED SHORTEST PATH MOVEMENT MODEL (TBW-SPMM)

Nodes in Shortest Path Movement Model (SPMM) use the Dijkstra's algorithm (Chen et al. 2003) to find out the shortest path from the node current location to the chosen destination on the map. The chosen destinations on the map are named Points of Interests (PoIs), where PoIs could be a restaurant, a shopping area,... etc. Weighted Shortest Path (W-SPMM) Movement Model is an improved version of SPMM where PoIs are visited with unequal probability. This is due to the fact that some places are more popular than others. This model is easy to understand and to be adopted in simulations, but does not assure inter-contact time and contact time duration that match real world traces when small number of nodes are used in the simulation (Ekman et al. 2008).

This raises the need to improve W-SPMM. This paper presents Time Based Weighted Shortest Path Movement Model (TBW-SPMM). The main idea behind the proposed model is some observation from human daily life activates. Examples of these observations are given below.

The four seasons: The probability of visiting group of PoIs during winter is different than their visiting probability during summer. For example, one may visit his preferable skiing resort during winter every weekend, but he will never do during summer since it is closed. This can be generalized to include the activities in four seasons (Spring, Summer, Fall, Winter) are performed with different probability.

The school semester: Students tends to visit PoIs, like registration office, with high probability at the first two weeks of their academic semester.

Whereas this high probability might goes to exam halls during the seventh and eighth week due to the fact that the 7th and 8th weeks of the academic semester are usually exam period.

The week activities: The week activates can be classified to weekdays and weekends activities. A PoI might receive a high visiting probability during weekdays, such as a café near business offices. However, this probability is expected to be reduced during weekends since nodes tend to travel differently to explore new areas during weekends.

The one day activities: A breakfast restaurant might have a high visiting probability as a PoI during morning hours, however, this probability might go to zero during evening hours. Thus, this probability reasonably might transfer to a dinner restaurant during evening hours.

These are not only the incidents where a clear modification of the PoI visiting probability is differing based on the time or period of the visit. This encourages the proposal of TBW-SPMM, where nodes move not only based on predefined probability for each PoI, but the period or the time of the visit is considered. For example, the 35% probability of visiting PoI A during winter season might be reduced to 5% due to the fact that A is an outdoor gym, and nodes prefer indoor gym during winter.

4 PERFORMANCE RESULTS

ONE (Keränen et al. 2009) is a discrete event simulation package. It combines movement modeling, routing, visualization and reporting. Mobility models determine node movement within the simulation environment. The Random Waypoint Model (RWP) is widely used and is based on random directions and speeds. However, this random node movement is unrealistic when mobile devices are carried by humans. It is more pragmatic to assume that nodes move towards a specific destination, then another destination, and so on. These destinations are typically particular locations such as malls, restaurants or schools, and so are called Points of Interest (PoI). The more realistic Shortest Path Movement Model (SPMM) has nodes moving towards particular locations. However, previous works show that SPMM did not match real-world traces when a small number of nodes are simulated (Ekman et al. 2008). Thus, the proposed TBW-SPMM is also simulated in ONE to be compared against real world experiment traces. The two movement models RWP and SPMM are simulated here for comparison purpose.

The simulation parameters employed are from the realistic environment described in (Altamimi

et al. 2017). The Helsinki City Scenario (HCS) model is used. The scenario has nodes moving in a part of the downtown Helsinki area. With HCS, node mobility is based on simulating 50 mobile users moving by foot, 30 by car, and 6 by trams in the streets of downtown Helsinki. Additionally, three sensor networks are randomly deployed in three locations; each network consists of 10 nodes. Each node represents a user moving with realistic speed along the shortest paths between different Points of Interest (POIs) and random locations. The trams follow real tram routes in Helsinki. The simulation area is $4500 \times 3400 \text{ m}^2$ size. A percentage of the nodes are assumed to visit each PoI. This percentage is defined with respect to the other PoIs locations. For example, a PoI has a 10% visit percentage if it is visited 90% less frequently than the location visited most frequently. The 30 location percentages are uniformly distributed from the set $\{10\%, 20\%, \dots, 100\%\}$. Each node represents a user moving with a realistic speed along the shortest path between locations randomly chosen based on their visit percentages. The visiting percentage

is changed every hour of the simulation period in TBW-SPMM.

The simulation environment parameters are summarized in Table 1.

The real world experiment has 128 devices as in the above simulated environment, one is a stationary node, 8 are monitored mobile nodes, and the remaining devices are external devices. The experiment was conducted in Intel Research Cambridge Corporate Laboratory (Chaintreau et al. 2006). This experiment has been widely used in literature (Ekman et al. 2008, Hossmann et al. 2011, Petz et al. 2009) to be a reference for real world trace, thus it is employed here.

To validate the realism of the proposed movement model, inter-contact time and contact duration time have been examined in the WPM, SPMM, TBW-SPMM against the real world experiment traces. Figure 1 shows that the proposed TBW-SPMM is the closest to the real world trace when the inter-contact time is employed for comparison. The figure in particular shows the number of nodes who had inter-contact less than T. Where T is the time (second), reported in x-axis.

Figure 2 show the total contact duration between nodes in the three models compared to the real world trace. The proposed TBW-SPMM again shows that it is more realistic than the other two models, namely RWP and SPMM since TBW-SPMM have similar contact duration between its nodes to the real world trace.

Table 1. Simulation environment parameters.

Parameter	Value
Transmit Rate	250 Kbps
Transmit Range	50 m
Message Size	50–150 KB

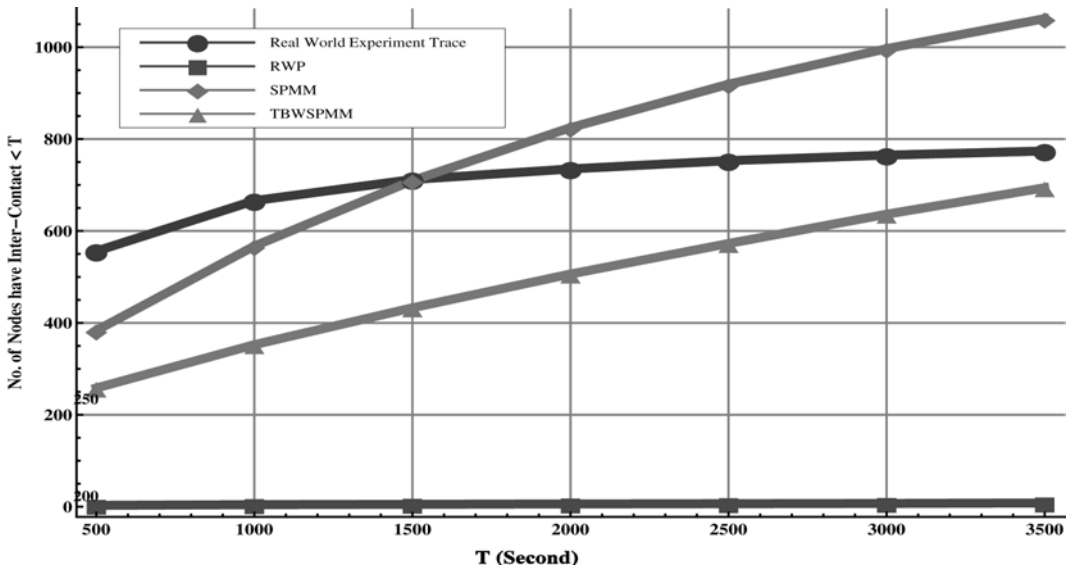


Figure 1. Number of Nodes that have inter-contact < Time (T).

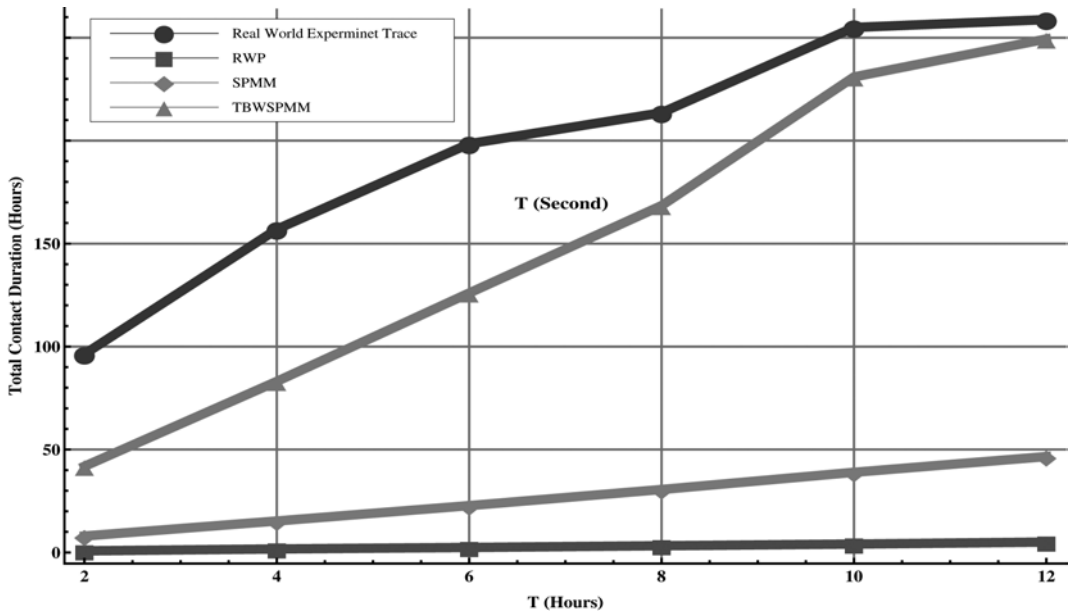


Figure 2. Total contact duration between nodes.

5 CONCLUSION

Movement models have been discussed in this paper. The main goal is to propose a movement model that can be as close to real world movement experiment as possible. Comparing the real world traces to previously proposed movement model shows that these model needs improvement to reach realism. Based on real world observation, TBW-SPMM is proposed. It is also validated to be close to real world experiment traces.

REFERENCES

- [1] Johnson, D.B., and Maltz, D.A. (1996) Dynamic source routing in ad hoc wireless networks. In *Mobile Computing, Imielinski and Korth*, Eds., vol. 353. Kluwer Academic Publishers.
- [2] Keranen, A., and Ott, J. (2007) Increasing reality for DTN protocol simulations. Technical Report, *Helsinki University of Technology, Networking Laboratory*.
- [3] Shahzamal M., Pervez M., Zaman M. and Hossain M. (2014) Mobility models for delay tolerant network: A Survey. *International Journal of Wireless & Mobile Networks (IJWMN)* Vol. 6, No. 4.
- [4] Walker, Brenton D., Clancy T., and Glenn K. (2008) "Using localized random walks to model delay-tolerant networks." In *Military Communications Conference. MILCOM*, pp. 1–7.
- [5] Petz, Agoston, Enderle J., and Julien K. (2009) "A framework for evaluating dtn mobility models." In *Proceedings of the 2nd International Conference on Simulation Tools and Techniques*, p. 94–99.
- [6] Zaninetti, L., and Mario F. (2008) "On the truncated Pareto distribution with applications." *Central European Journal of Physics*, no. 1, pp. 1–6.
- [7] Ekman, F., Keränen, A., Karvo J. and Ott, J. (2008) Working day movement model, in *ACM WMM*, 33–40.
- [8] Keränen, Ari, Teemu Kärkkäinen, and Jörg Ott. (2010) "Simulating Mobility and DTNs with the ONE." *Journal of Communications*, no. 2, pp. 92–105.
- [9] Chen, Jing-Chao. (2003) "Dijkstra's shortest path algorithm." *Journal of Formalized Mathematics*.
- [10] Keranen A., Ott J., and Kärkkäinen T., (2009) "The ONE simulator for DTN protocol evaluation," in *Proc. Int. Conf. on Simulation Tools and Techniques*, Rome, Italy, pp. 1–10.
- [11] Altamimi, Ahmed B., (2017) Buffer Management in Mobile Social Networks. In *Journal of Computers*. Vol.12(1): 20–27.
- [12] Chaintreau, A., Hui, P., Crowcroft, J., Diot, C., Gass, R., and Scott, J. (2006) Impact of human mobility on the design of opportunistic forwarding algorithms. In *Proc. INFOCOM'06*.
- [13] Hossmann, Theus, Thrasyvoulos Spyropoulos, and Franck Legendre. (2011) "Putting contacts into context: Mobility modeling beyond inter-contact times." In *Proceedings of the Twelfth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pp. 18–22.

Video steganography using P-frame motion vector

A.E. Ibrahim & M.A. Elshahed

Department of Physics, Faculty of Women for Arts, Sciences and Education, Ain Shams University, Cairo, Egypt

T.I. Elarif

Department of Computer Science, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt

ABSTRACT: Steganography is the art and science of writing hidden messages in such a way that no one apart from the sender and intended recipient even realizes there is a hidden message. Video Steganography algorithm was applied in this paper; data was embedded in blocks of P frames with larger magnitude. The larger magnitude refers to the faster moving speed of the macro-blocks. In this case, the distortion results from data embedding is minimum compared to the distortion result from modifying all motion vectors. From the results, we found that at constant threshold, the less the number of the selected blocks, the large the number of frames needed for the embedding process, the quality of the stego frame increases and low degradation is obtained.

1 INTRODUCTION

The rapid growth of multimedia and internet allows for wide distribution of digital media data. It becomes much easier to edit, modify and duplicate digital information. Also, digital documents are easy to copy and distribute, so it will be faced by many threats. To solve this problem steganography technique is used.

The term steganography comes from the Greek words *stegano* (cover) and *graphy* (write). As result a steganography literally means covered writing.

The basic model of steganography consists of Carrier, Message and Password, as shown in Figure (1). Carrier is also known as cover-object, in which the message is embedded. Message is the data that the sender wishes to remain confidential such as plain text, cipher text and other images. Password is known as stego-key, which ensures that only the recipient who knows the correspond-

ing decoding key will be able to extract the message from a cover-object. The cover-object with the secretly embedded message is then called the stego-object.

The Steganographic systems use multimedia objects like image, audio, video etc as cover media. Depending on the type of the cover object, a suitable technique is selected for the embedding process (Kh. Al-Ani, Zaidoon et al., 2010) (Chakraborty & Kumar Bandyopadhyay, 2012) (Amin et al., 2003).

There are mainly four requirements of any information hiding technique, namely, Imperceptibility, Capacity, Security and Robustness. Imperceptibility means that human eyes cannot distinguish the difference between the stego frame and the original frame. Capacity refers to the amount of data that can be embedded in the cover object. Security means that the embedded information can't be removed after the attacker discovers it. Robustness means that the ability of the embedded data

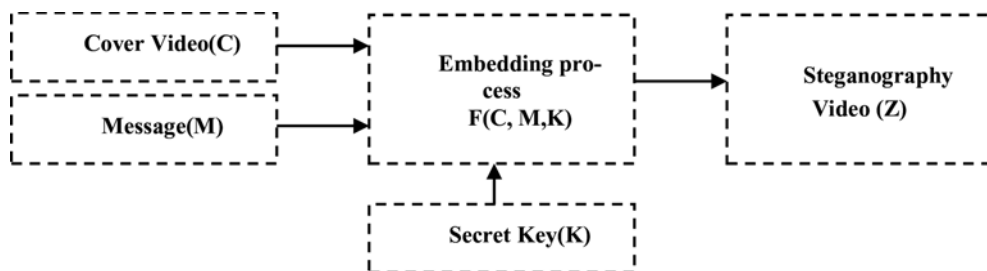


Figure 1. Basic steganography model.

to remain intact if the stego-system undergoes transformation such as linear and non-linear filtering; addition of random noise; and scaling, rotation, and loose compression (Zaidoon et al., 2010) (Muttoo & Kumar, 2009).

Video Steganography is a technique to hide any kind of secret information into a carrying video file. The use of the video based steganography can be more eligible than other multimedia files, because of its size and memory requirements (Swathi & Jilani, 2012).

Compression is the basic process of reducing the size of data in order to obtain efficient storage and transmission. There are two compression techniques known as lossless and lossy techniques. In lossless the reconstructed image after compression is identical to the original image. In lossy the reconstructed image is not identical to the original image because some of the original information is lost. Lossy method is capable of achieving a high compression compared to the lossless method.

Video compression plays an important role in real-time video conferencing applications. Video compression techniques are about removing and reducing redundant video data (removing temporal and spatial redundancies) so that a video file can be easily sent over an internet and stored on removal or computer disks (N. Manjanaik. & R. Manjunath., 2013) (Shaikh & Badnerkar, 2014).

MPEG (Moving Picture Experts Group) is a widely used and popular video compression standard. In a MPEG, any video is considered as a sequence of images. Group of Pictures (GoP) is then defined as a set of video frames that follow a predefined frame pattern. Frame pattern consists of Intra frames (I), Bi-directional frames (B) and Prediction frames (P). I-frames are used as reference/anchor frame to predict P/B-frames in a GoP. P-frames are predicted from preceding I-frame/P-frame. B-frames are predicted using preceding and following I/P- frames within a GoP. B-frames can be optional in a video sequence (Jaiswal & Dhavale, 2013).

In MPEG video sequences, most frames are encoded using motion compensation prediction. In P frames each macro-block has one motion vector, while in B frames each macro-block has two motion vectors.

We used the peak signal to noise ratio (PSNR) and correlation to measure the quality of the reconstructed frame. The PSNR is defined as:

$$PSNR = 10 \times \log_{10} \left(\frac{MAX_I^2}{MSE} \right)$$

where the Mean Squared Error (MSE) is defined as:

$$MSE = \frac{1}{mn} \sum_{i=0}^m \sum_{j=0}^n [I(i, j) - K(i, j)]^2$$

where I is the original frame and K is the modified (reconstructed) frame. MAXI is the maximum possible pixel value of the frame, e.g. the MAXI value is 255 for frames with 8-bit colour depth (Liu et al., 2011).

2 PREVIOUS WORK

In (Changyong et al., 2006) a steganographic algorithm in MPEG compressed video stream was proposed. In each GOP, the control information to facilitate data extraction was embedded in I frame, in P frames and B frames, the actually transmitted data were repeatedly embedded in motion vectors of macro-blocks that have larger moving speed, to resist video processing.

In (Sherly & Amritha, 2010) a new Compressed Video Steganographic scheme was proposed. In this algorithm, data hiding operations are executed entirely in the compressed domain. data are embedded in the macro blocks of I frame with maximum scene change and in block of P and B frames with maximum magnitude of motion vectors. To enlarge the capacity of the hidden secret information and to provide an imperceptible stego-image for human vision, a novel steganographic approach called Tri-way Pixel-Value Differencing (TPVD) is used for embedding.

In (Pan et al., 2010) a novel video steganography scheme based on motion vectors and linear block codes was proposed. In this paper secret messages embed in the motion vectors of cover media during the process of H.264 compressing. Linear block codes has been used to reducing the modification rate of the motion vectors.

3 STEGANOGRAPHIC ALGORITHM

3.1 Data embedding algorithm

In our algorithm, the data were not embedded in each motion vector of P frames, but in the motion vectors which have larger magnitude than certain threshold. The details of data embedding are shown in Figure (2).

1. For a P frame, get the motion vector $PMV[i]$, $0 < i < NMB$ from the compressed video stream.
where NMB: is the number of macro-blocks in this frame.
2. Calculate the magnitude of motion vector

$$|PMV[i]| = \sqrt{H^2[i] + V^2[i]}, \quad (1)$$

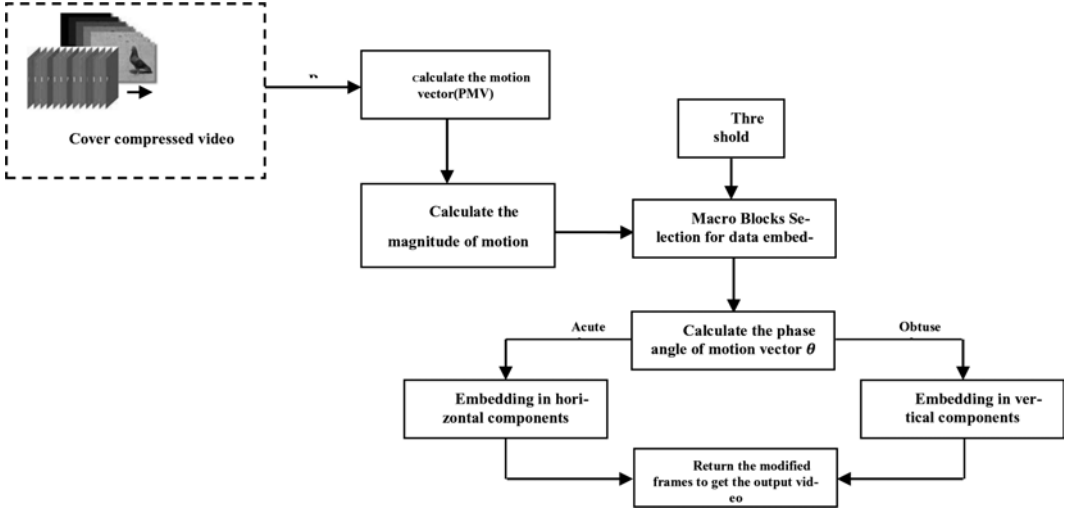


Figure 2. Block diagram of the proposed algorithm.

where

$H[i]$: is the horizontal component of motion vector in the i th macro-block.

$V[i]$: is the vertical component.

3. If the threshold of the magnitude of motion vector is ϵ , we can select the embeddable macro-blocks.

$$MB[i] = \begin{cases} 1 & |PMV[i]| > \epsilon \\ 0 & |PMV[i]| \leq \epsilon \end{cases}$$

where, $MB[i] = 1$: denotes the macro-block that satisfies the condition, and can be used for embedding.

$MB[i] = 0$: denotes no data embedding in this macro-block.

4. Calculate the phase angle of motion vector

$$\theta[i] = \arctan\left(\frac{V[i]}{H[i]}\right) \quad (2)$$

5. Embed data in the selected macro-blocks, based on the value of θ .
 - a. If θ is acute angle, less distortion will be introduced by modifying horizontal component of motion vector, so the data were embedded into the horizontal component.
 - b. If θ is obtuse angle, less distortion were introduced by modifying vertical component of motion vector, so the data were embedded into the vertical component.
6. The same calculations were repeated for all the P frames in a GoP, we do, and the number of

eligible motion vectors were achieved. For each eligible motion vector, one bit can be embedded.

3.2 Data extraction algorithm

To extract the hidden data, we follow the same steps used in the data embedding algorithm, get the motion vector $PMV[i]$, $0 < i < N_{MB}$. Where N_{MB} : is the number of macro-blocks in this frame, calculate the magnitude of motion vector using equation (1).

Depending on the threshold of the magnitude of motion vector which was used in data embedding, we can select the embeddable macro-blocks, calculate the phase angle (θ) of motion vector using equation (2). If θ is acute angle, data was embedded into horizontal component, then data extracted. Also, if θ is obtuse angle, data was embedded into vertical component, then data extracted. For all used P frames, the above calculations were repeated until all embedded data was extracted.

4 EXPERIMENTAL RESULTS

In our algorithm we used a dataset that has 300 frames, the frame size is 288×352 , the macro-block size 8×8 pixels, the motion search window is set to 16×16 pixels, the length of a GoP is set to 15 and a text message “A computer network is the infrastructure that allows two or more computers to communicate with each other” was used as a secret message.

In the experiment, at constant threshold = 7 (To select a suitable threshold, we applied an experiment studying the relation between the threshold and

the PSNR and we found that threshold = 7 is suitable for this work) we change in the number of the selected blocks (change in the embedding capacity of the frame at constant threshold). We study the relation between the number of selected blocks and the number of P frames used in the embedding process, and the relation between the number of selected blocks and the PSNR.

Figure (3) shows the PSNR values of the stego frames after data embedding at different numbers of the selected blocks. From the results we found that at number of blocks = 20, the average PSNR of the stego frame = 51.35056257, number of GoP selected for data embedding = 14. At number of blocks = 30, the average PSNR of the stego frame = 49.33886273, number of GoP selected for data embedding = 9. At number of blocks = 40, the average PSNR of the stego frame = 46.3784794, number of GoP selected for data embedding = 6.

At number of blocks = 50, the average PSNR of the stego frame = 45.55797157, number of GoP selected for data embedding = 5. Without limiting the number of blocks (for all the available embedding capacity at threshold = 7) the average PSNR of the stego frame = 42.35961037, number of GoP selected for data embedding = 2. From this we conclude that the less the number of the selected blocks, the large the number of frames needed for the embedding process, the quality of the stego frame increases and low degradation is obtained.

Figure (4) shows the correlation values of the stego frames after data embedding at different numbers of the selected blocks. We found that when the number of selected blocks decreases, the correlation of the stego frames increases.

Figure (5) shows the PSNR values of the stego frames after data extraction at different numbers of the selected blocks. From the results we found

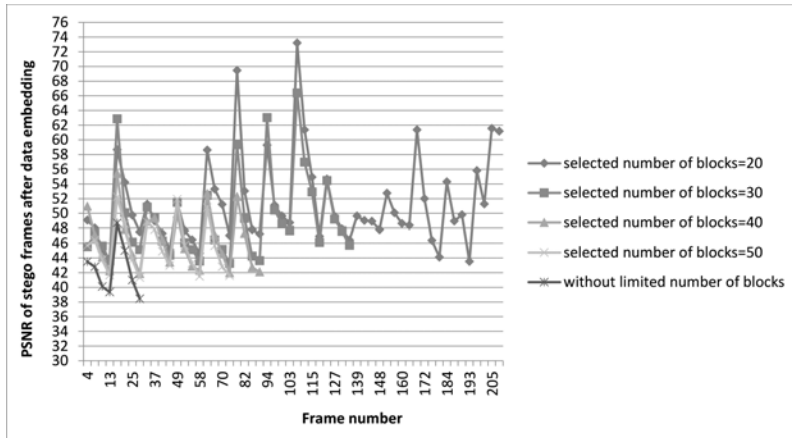


Figure 3. The relation between the frame number and PSNR of stego frames after data embedding.

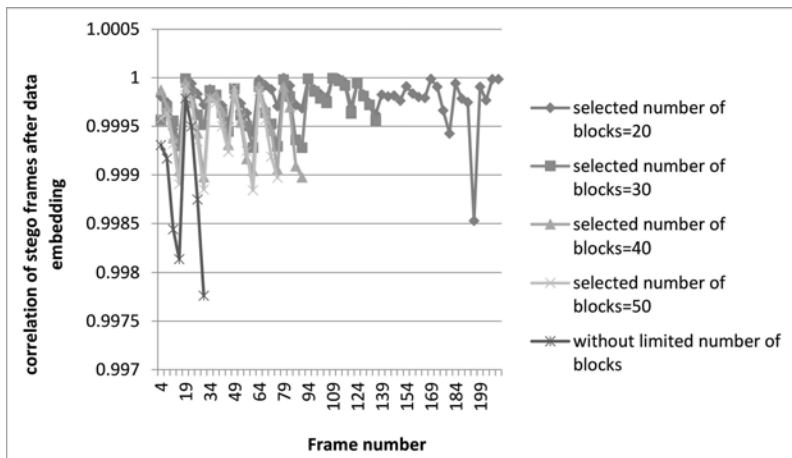


Figure 4. The relation between the frame number and the correlation of stego frames after data embedding.

that at number of blocks = 20, the average PSNR of the stego frame = 50.98155209. At number of blocks = 30, the average PSNR of the stego frame = 48.85195449. At number of blocks = 40, the average PSNR of the stego frame = 45.91406377. At number of blocks = 50, the average PSNR of the stego frame = 44.88980126. Without limit the number of blocks (for all the available embedding

capacity at threshold = 7) the average PSNR of the stego frame = 41.99672931. From this we conclude that by decreasing the number of selected blocks, the average PSNR increases, and low degradation is obtained.

Figure (6) shows the correlation values of the stego frames after data extraction at different numbers of the selected blocks. Figure (7) shows

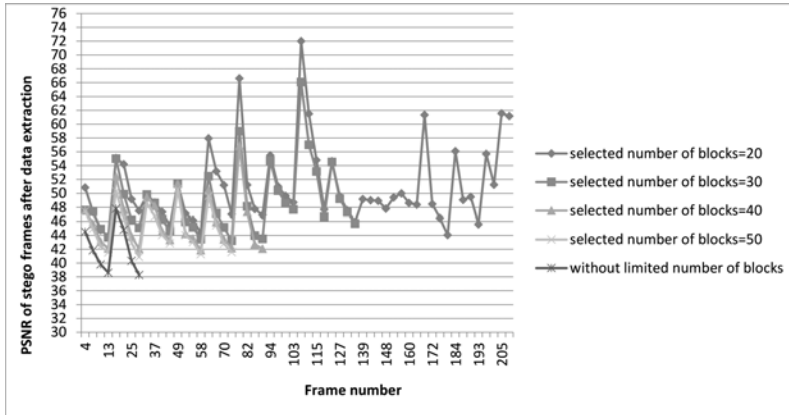


Figure 5. The relation between the frame number and the PSNR of stego frames after data extraction.

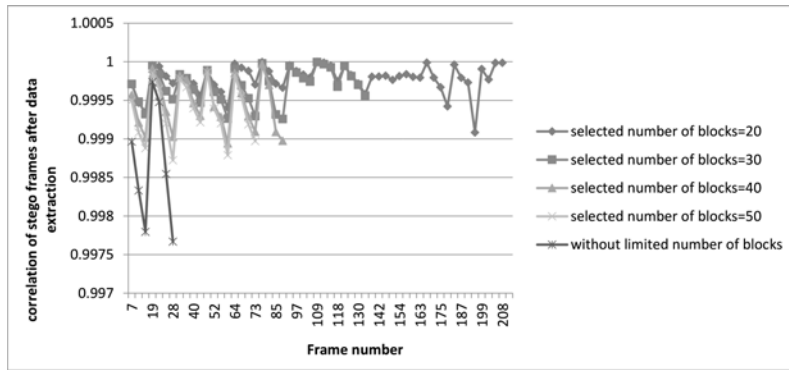


Figure 6. The relation between the frame number and the correlation of stego frames after data extraction.

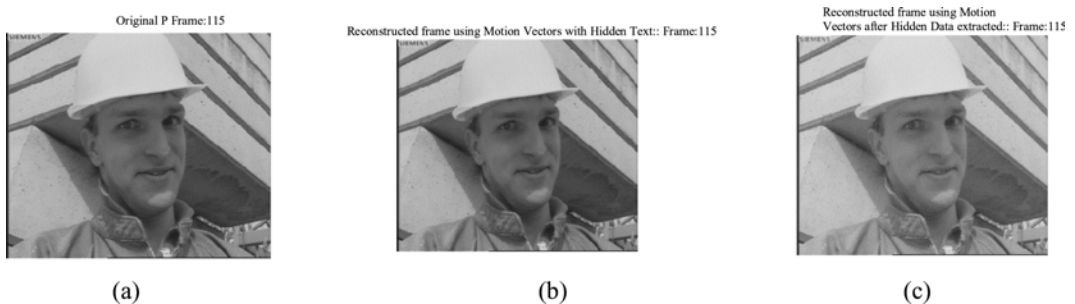


Figure 7. Illustrates selected frame from the dataset, (a) the original P frame, (b) the reconstructed frame using motion vectors with hidden text, (c) the reconstructed frame using motion vectors after hidden text extracted.

an example for selected frame used for embedding in dataset.

5 CONCLUSION

We applied a video steganography algorithm by using P frame motion vector. At constant threshold, we study the relation between the number of selected blocks and the average PSNR of stego frames and the relation between the number of selected blocks and the number of P frames used in the embedding process. From the results we found that, the less the number of the selected blocks, the large the number of frames needed for the embedding process, the quality of the stego frame increases and low degradation is obtained.

REFERENCES

- Al-Ani, Zaidoon Kh., Zaidan, A.A., Zaidan, B.B. & Alanazi, Hamdan. O. March 2010, Overview: Main Fundamentals For Steganography, *Journal Of Computing*, Volume 2, Issue 3, Issn 2151-9617.
- Amin, Muhalim Mohamed, Ibrahim, Subariah, Salleh, Mazleena & Katmin, Mohd Rozi. 2003, Information Hiding Using Steganography, *Universiti Teknologi Malaysia*.
- Chakraborty, Suman & Kumar Bandyopadhyay, Samir. August 2012, Two Stages Data-Image Steganography Using DNA Sequence, *International Journal Of Engineering Research And Development* E-Issn: 2278-067x, P-Issn: 2278-800x, Volume 2, Issue 7, Pp. 69-72 69.
- Changyong, Xu., Xijian, Ping & Tao, Zhang. 2006, Steganography in Compressed Video Stream, *Proceedings of the First International Conference on Innovative Computing, Information and Control (ICICIC'06)*.
- Jaiswal, Sunil & Dhavale, Sunita. 2013, Video Forensics in Temporal Domain using Machine Learning Techniques, *I. J. Computer Network and Information Security*, 58-67.
- Liu, Zhao., Qiao, Yuansong, Lee, Brian, Fallon, Enda, Karunakar, A.K., Zhang, Chunrong & Zhang, Shuaijun. 2011, Experimental Evaluation of H.264/Multi-view Video Coding over IP Networks.
- Manjanaik, N. & Manjunath, R. 2013, Selection Of Intra Prediction Modes For Intra Frame Coding In Advanced Video Coding Standard, *Ijret: International Journal Of Research In Engineering And Technology*, Eissn: 2319-1163 | Pissn: 2321-7308, Volume: 02 Issue: 12.
- Muttoo, S.K. & Kumar, Sushil. 2009, Robust Source Coding Steganographic Technique Using Wavelet Transforms, *BVICAM's International Journal of Information Technology*, Vol. 1 No. 2 ISSN 0973-5658.
- Pan, Feng, Xiang, Li, Yang, Xiao-Yuan & Yao Guo. 2010, Video Steganography using Motion Vector and Linear Block Codes.
- Shaikh, Muhammad Aakif & Badnerkar, Sagar S. March 2014, Video Compression Algorithm Using Motion Compensation Technique: A Survey, *International Journal of Advance Research in Computer Science and Management Studies*, ISSN: 2321-7782, Volume 2, Issue 3.
- Sherly, A.P. & Amritha, P.P. August 2010, A Compressed Video Steganography using TPVD, *International Journal of Database Management Systems (IJDMS)*, Vol. 2, No. 3.
- Swathi, A. & Jilani, S.A.K. 2012, Video Steganography by LSB Substitution Using Different Polynomial Equations, *International Journal Of Computational Engineering Research (ijceronline.com)*, Vol. 2, Issue 5.

A study to evaluate the role of social networks in supporting the educational process: Survey study

Ali Almajhaddi & Saad Almutairi

King Abdul-Aziz University, Jeddah, Saudi Arabia

ABSTRACT: Social networks play an important role in the generation and transfer of knowledge from one generation to another within communities if you use a way that promotes learning opportunities and expand the horizon for the exchange of knowledge among users which will reflect positively on the academic achievement of the social networks to overcome the barriers to place the material and facilitate convergence between the parties to the process level making educational knowledge generation and passed on is easier. In this paper, we will seek to know the role played by social networks in the generation of knowledge and its role in raising the academic achievement level by offering some important in various states of case studies in both developing and developed countries, especially the Arab countries with similar circumstances with the nature of Saudi society, and thus benefit of it in Saudi Arabia environment. Also, we are targeting to be informed about important points concerning the following: knowledge of the reality of knowledge generation process and the extent of interest in this process; know the nature of the relationship between the generation of knowledge through social networks and their impact on academic achievement, and finally the detection of anomalies in the method used to obtain knowledge, generated through social networks.

Keywords: knowledge acquisition, knowledge creation, organization, academic affairs, community and deicing

1 INTRODUCTION

There is no doubt that social networking an effective role in strengthening learning processes, only evidence of modern psychology, the process of storing the human mind to the information, or vocabulary, is determined by its nature the psychological state of the recipient, and then the need for a recreational dimension during the lecture process, this made possible by social networking sites, where students will be more motivated, especially when learning languages, mathematics, social and material, it is also active skills of learners, and provide opportunities for them, and encourage their creative thinking patterns and different ways, and the pride of the positive role of the learner in the dialogue, and make it an active participant with the others, and promote educational methods in a collaborative environment, and helps the learner to studying constructive by offering a diverse and integrated training, and allows the teacher and the learner the possibility of exchanging books and provide educational games aimed at making use of opinion polls, where the teacher uses these surveys as an educational tool effective and increase communication with students and follow-up of new

news and stand on what informed of current events and political, economic, scientific and socially [1].

It also helps to instill ambition in the hearts of the learners by encouraging them to create and design new applications to networks serving the educational material, and disseminated among the educated to take advantage of them, where many of the students presented their applications process among themselves, such as the number of schools and international institutes students who make up the groups on the site, and contribute to the transfer of education from the competition phase, the integration stage, by requiring all learners to participate in the dialogue and to gather information, [2] and then make teaching and learning more fun and energetic and experience the clock, and the participation of the challenge, where the teacher can engage his students in the implementation of projects relate to promote their institutions teaching, in order to measure their talents and enrich their abilities, and the extent of their self-confidence, and the introduction of new methods, are encouraged to put forward ideas, and promote a spirit of partnership and communication between learners, and to enable the teacher to put himself Office hours face hours, allows the students through which to

communicate with him and ask questions and receive the answers [3]. The social networks, check the great benefit of education, where he was able to access the various portable devices, and facilitate the distribution of scientific material process in the classroom, as well as facilitate the evaluation process, and testing. So we can say there has been various overview and opinions which recognized four major advantages of social media use in higher education. These include, enhancing relationship, improving learning motivation, offering personalized course material, and developing collaborative abilities. This means that social networking activities have the possibility of enhancing student contact and is used to improve their participation in class, particularly where introverted students are involved. Students can function in online group learning, with less or no anxiety of needing to raise questions before peers at school.

2 BACKGROUND AND PRELIMINARIES

The social networks one of the most important media, which escalated their star in cyberspace, despite its modernity, demand for them has doubled, and became a play roles influential, politically, economically and socially, and their influence extended to the field of education, where he sees is one senior education experts, she added aside from the human form, through the participation and interaction of the human element in the educational process, helping to increase the desire for education, although this does not preclude the presence of some of the drawbacks.

Historically featured social networking services, before the demise of the second millennium the sun, as a component Chairman of the techniques of the second generation of the Web, targeting more communication and informational exchange. In terms of numbers, numbers of them doubled account on these networks, and was the director of marketing at Facebook, [5] it has said that the number of users of the site, bypassing the time being the 900 million people around the world, and that during the past year, the turnout was steady in the Middle East and North Africa region, and in the field of education has emerged the role of social networks dramatically during the past few years, where Subscribe where thousands of schools, colleges and universities around the world, not to mention the involvement of students for the purpose of education, which have benefited from its services in the framework of the so-called hybrid education, cultural or educational system. In order for there continues to be an educational building, the real advantage of the social networking services, it is both the teacher and the

learner must access according to a set of requirements and regulatory considerations, we mention the most important:

- Before embarking on the course, the teacher can establish a page on any of the networking sites with the participation of experts and interested students, and will take their opinions, which helps him to determine the content and formulation of stated objectives.
- Conduct interactive discussions On time discussions, about the important issues.
- Students are divided into groups in the event of collective tasks such as graduation projects.
- Send messages to an individual or a group of students through the profile when needed.
- Delivery and receipt of homework and other school functions.
- Could use some social networking tools, such as face book icons or comment or like to take the students' opinions about the components of the subject matter.
- Select the category that will benefit from the learning process precisely.
- Create a page (Page) or group (Group) closed with a membership of only beneficiary category, with the possibility of control to add or not to add new members from outside.
- A clear definition of the goals of the group and purpose.
- The appointment of the commander of the group, a faculty member who can be a student appointed as secretary of the group.
- The definition of the principles and behaviors of the group and the organization of educational process.

And the application of these requirements and regulatory considerations, can reap a lot of advantages and benefits of the services offered by the education systems across the social networks, where can the group's leader (faculty member) offer educational substance on his students, and participation by raising educational issues, and engage in a constructive about each lesson discussion the lessons of art in the courtyard of the dialogue, and can also be put to his students specific assignments, and then ask them to search for them and resubmit them, so that they can stand on what each of them reached separately, and put appropriate assessment, and can offer them a problem, and asked to put every one of them in response to that problem in a private message, and have the added images and sound clips and video related article or one of her classes, including enriches article or lesson, and helps to understand better, and be either produced by the teacher or the learner [6], with the possibility of post and add links to pages on the Internet, offering the further enrichment of

the material educational and discuss the content, and determine prior appointment to meet it with his students at the same time, to respond to any immediate inquiry, or dialogue and debate about a subject, and take advantage of the chat on social networks, discussion of some elements lesson between the teacher and the learners or some of them, or between the learners themselves, and create new applications that will enrich the material and its lessons.

3 LITERATURE REVIEW

Social media has proof its ability to boost the communication between people and many industries are attempting capitalized on the power of social media. But, there is one industry in particular that is best suited to adapt to these new mediums— institutions of higher education. As social networking has become one of the most popular means of communication among the traditional college-age demographic universities are beginning to utilize these technologies to communicate with current and prospective students [7] The Educes Center for Applied Research (ECAR) Study of Undergraduate Students and Information Technology, 2008 defines social networking sites as “Web-based services that allow construct a public or semi-public profile within a bounded system, 2) articulate a list of other users with whom they share a connection, and 3) view and traverse their list of connections and those made by others within the system According to [8], some site published the survey result of social media usage. Socialnetworkingwatch.com states that 68% women and the rest are men. Socialnomics.net mentioned some fact which show the grow of social media, such as: Years to reach 50 million users: radio (38 years), TV (13 years), internet (4 years), iPod (3 years), and Facebook (100 million users in less than 9 month). If Facebook were a country it would be the world’s 4th largest between the United States and Indonesia (note that Facebook is now creeping up—recently announced 300 million users) Social media take on many different forms, such as internet forums, weblogs, social blogs, Micro blogging, wiki, podcast, photo or video sharing, rating and social bookmarking. [9] divide the social media types into six categories such as (Wikipedia):

- Collaborative project (ex. Wikipedia)
- Blogs and micro blogs (twitter)
- Content communities (YouTube)
- Social networking sites (Facebook)
- Virtual game worlds (world of war craft)
- Virtual social worlds (second life)

Technologies used in social media include: blog, picture-sharing, blogs, wall-posting, email, instant messaging, music-sharing, crowdsourcing, and VoIP. Functionality and the usage of Social Media Network has spread into several aspects as described in frame work proposed by [9]. This frame work show the seven building block of social media functionality, which are: identity, conversation, sharing, presence, relationship, reputation and groups.

4 OBJECTIVES OF THE STUDY: THE AIM OF THIS PAPER IS TO

1. knowledge of the role played by social networks in the generation of knowledge between different generations, and the reality of the process of knowledge generation and the extent of interest in this process within Saudi society.
2. Assess the role which played by social networks in raising the level of academic achievement by providing some important countries in the various case studies in both developing and developed countries, especially the Arab countries with similar circumstances with the nature of Saudi society, even benefit from it in Saudi Arabia environment.
3. Know the nature of the relationship between the generation of knowledge through social networks and their impact on academic achievement for the users of social networks within Saudi society.
4. Detect anomalies in the method used to acquire knowledge, which was born through social networks.

5 METHODOLOGY

Review of previous studies that looked systematically at issue with respect to evaluation the role of social networks in promoting education operations, has also been conducting a search Scholar Google. Search for words and phrases included Facebook, Education, Higher Education, social media, and social media in education, MySpace, LinkedIn, Web 2.0, social, and social networking Web sites and blogs, and the use of the most recent business to find the relevant materials to the fore their materials additional. Since keywords in articles and research on joint lists are not based, as the researcher used the approach to comparative research when conducting surveys and analytical research methodology to analyze the role of networking in education operations in both developing and developed countries and the Arab countries. Research investigators have recognized that social

network sites are a rich source of behavioral data. With permissions in place, physically and digitally, researchers can sift and sort through a myriad of posts, pokes, and tweets in order to examine such variables as content, attitudes, and understanding. In some cases, automated, collection techniques can capture large datasets containing profile updates, linkages, and usage trends which are then able to be explored.

6 KNOWLEDGE GENERATION AND SOCIAL NETWORKS

The knowledge generation respect to processes that focus on access, purchase, innovation, and discovery, and the acquisition and the acquisition of knowledge that can generate knowledge through a number of processes that extend between the challenge of creativity and the serious research, meetings informal, seminars and all the ways the crisis so. Usually generate existing institutions and use of knowledge through the conversion process known as means transforming the knowledge and tacit knowledge to know the undeclared, and vice versa, it is believed that the organization may not generate the knowledge of its own, because the tacit knowledge held by individuals is the regulatory basis of knowledge generation process [10]. Thus, should the accumulation of tacit knowledge that has been collected at the individual level, then the organization is expanded through the four modes of transfer of knowledge on the basis of the following patterns process:

- Society: converting tacit knowledge to tacit knowledge
- Anthropomorphism: converting tacit knowledge to know the unspoken
- Composition: transforming knowledge declared that knows the unspoken
- Assimilation: diversion of declared tacit knowledge knowledge

This means knowledge creation, as the process by which this generation of knowledge through the participation of teams and support working groups to generate the capital of the new knowledge-based in the new issues and practices that contribute to the definition of the problems and find new solutions in innovative continuous, as the organization that provides the ability to excel in achievement and achieve high market position in the different areas such as exercise and start a new strategy for action and speed up the solution of problems and practices and the development of professional skills and assist management in the recruitment and retain talent pipelines. This reinforces the need to understand that knowledge and innovation

double two-way process: knowledge is the source of innovation for innovation and then come back and become a source of new knowledge.

And you can repeat institutions access to knowledge in many ways depending on the knowledge you are looking for the type you might get them through documents, reports and applications to the organization or through the development of expertise networks are online. So finding workers experience in the organization that has the knowledge, and sometimes organizations create knowledge through data discovery, or the use of knowledge as engineer's stations models can new knowledge discovery, as well as can get knowledge from outside, such as reports, industry sources and legal opinions and statistics scientific research and government.

Social networks can be divided into two main sections

- * Own internal networks, these networks, and to be from a group of individuals who represent a closed society or, in particular, represents individuals within the company or pond or within the institution or educational organization and controls necessary to call these people and not others to enter the site and participate in the recording and exchange of views Activities look no laps and attend meetings and participate in direct discussions and other activities, such as the network of these networks is made up of a group of individuals who represent a closed society or individuals that are within a company or pool or within the institution or organization and controls to invite these people only and not for others to access the site and to participate in the audit activities and the exchange of views and files, and attend meetings and participate in direct discussions and other activities such as LinkedIn network
- * General and extranets and networks available to all Internet users, but it is specifically designed to attract users to the network and allows multiple users to participate in its activities as soon as the user time to register on the site and submit the same to the site, this network Facebook. And identifies social networking content that is created by using social networking sites and special tools "user-contributed content" that is created by individuals on sites that encourage the creation and sharing of content. Content ranging from text messages, which are exchanged photos and videos with my various notes. Even social networking and interactive networks allow users to connect at any time and from anywhere via technical means certain programs [11].

Various features of social networks can be stated as follows

- It depends on the programs and systems to transfer information controlled by the major powers in the world, which breaks the control of the barrier personal or collective, and turning it into a global international.
- It depends on where you are the communication process through the private addresses where an individual control and speech on behalf of the private corridor, and through fixed or portable electronic devices.
- and enable communication beyond the borders of the region or the state, or what is known as spatial barriers breakdown in communication.
- to create a virtual world that simulates the technical reality with data from modern high-tech.

7 ROLE OF SOCIAL NETWORKS IN EDUCATION

Apparently, social networks are being used increasingly by university students. It is promoting virtual communities and virtual learning environments (VLEs) for expanding distributed learning among users. The students interact in their virtual communities freely with members of the community. They can share information and study experiences, research projects and job opportunities with each other. Various factors contribute towards the use of social media for educational purposes.

Educational institutions that reflect the world in which we live, a world characterized by multi-level social interactions, but this is the duty of the educational institution that encourages students to engage in these social activities. We have to teach our students, and show them any power they can possess, that do good use of social networks, and to change their view of education, also change their view of themselves, and social lives, we are sadly, does not address the enormous benefits of the social interaction of education through social networking, but it is our duty to begin to change this, and adopt a more open educational policies [12].

Spin a lot of questions about the importance of social networking in education, and on top of these networks Facebook—MySpace These are some points that should stand out:

All the social networking constant, regardless of the type of computer or system used to run, or even a browser. It is wonderful to use unfamiliar tools large section of the students, the students will obey to complete special scientific their discussion, in class, what will enrich the scientific article. Some students may not have an account

on these networks, but they can establish a provisional account, There are some students cannot distinguish between what is academic and what is a personal interview.

The use of social networking in education led to the development in the educational process, as a positive effect on the way the performance of the teacher and the learner and their achievements in the classroom because they contain a variety of information in various fields, social The network electronic contributed to a large and positive role in the educational field, and between those roles the network in the educational field:-

- Tool to save the information.
- Contributed to the interest in individual or self-education.
- Develop informatics capabilities of the students.
- Develop scientific thinking skills.
- Help develop creative thinking.
- Check some learning goals.
- Help in finding strategies and plans to solve some educational problems.
- Facilitated for individuals contact between educational institutions of the world far between the parties.
- Through electronic social network, can be found on the fields, periodicals and journals, books, articles and miscellaneous reports.

That social networking is not just a site to meet new friends or communicate with friends, or find out what is going on around the world, it's also a learning tool impressive if it is used effectively an important resource for information, and teachers can use in the classroom, especially in higher education, in order to improve communication, and the integration of students in the effective activities differ from traditional teaching methods, this method also students and young people recognize the other uses a more useful and effective for Facebook. It could be argued that there are a lot of ideas that can benefit the university teacher in different disciplines to increase the effectiveness of teaching, and also to draw the attention of students in order to use Facebook in the areas of benefits accrue to them [14]. These ideas include the following: follow developments in the specialty: the teacher can cost students search for new developments in the field of scientific material studied, and thus maintains the link students with new information in the specialty. Review of books and research collaboratively: Students and teachers can review the research together by sending them to students in the same specialization for consultation, as well as the teacher, and the supply of feed back on Facebook.

Educational toys: can be used to improve reading skills, especially English as a second language, where these games will increase stocks of terminology in the English language among students. Polls: used by the teacher as an educational tool effective and also to increase communication between the course students on Facebook.

- English Language Teaching: where students can communicate with other native speakers of the English language through groups or networks.
- Find sources of private student information: especially journalism students, where they will be able practical application of their specialization, through the use of feed with Facebook to follow the political and sports breaking news and news updates universities center. Follow-up to the new news: follow through groups of new news on the global sites like News
- Weather or natural disasters or new in medicine and science, where there are plenty of sites on the network useful for students of medicine, engineering and science.
- Building Applications on Facebook: where many of the students presented their applications process it, like many of the world's university students, who make up groups on the site.

Social networking websites provide tools by which people can communicate, share information, and create new relationships. With the popularity of social networking websites on the rise, our social interaction is affected in multiple ways as we adapt to our increasingly technological world. The way web users interact and talk to each other has changed and continues to change. These users now socialize through the internet and it takes away from the person socialization that has been around forever. Social networking websites have affected our social interaction by changing the way we interact face-to-face, how we receive information, and the dynamics of our social groups and friendships [15].

8 VARIOUS CASE STUDIES OF SNS IN EDUCATION

Drew widespread use of these networks such as Facebook and Twitter and other attention of those in charge of educational institutions in the countries of the world, they represent a different and advanced from the traditional education suitable for the education environment, to proceed to an open education depends communication and participation mainly to the educational process as a substitute for indoctrination, and give a broad horizon for the exchange of experiences and other experiments can be used to raise the innovation

and creativity of the student. So the whole number of e-learning specialists that social networks represent an appropriate environment for the education of a modern and sophisticated and meets the needs of the individual, and in line with the global technological revolution. No wonder it has become a social networking prevalent among members of society, and contributed to the possibility of access to those networks even through mobile devices that have become widespread among all segments of society and the spread of Internet networks as well as easily used. It became part of their daily lives, so it was used as a means of education as something natural does not represent a burden on them, and mix fun with science to reach the desired knowledge. It has headed a number of countries in the world to adopt social networking and the way of the most important means of education, In the following we'll mention some of the important case studies that deal with application of SNS in education and the efficiency of this tool in improving the quality of education [16].

8.1 Case study one: Chinese experience

A lot of Chinese schools, in the capitals of Chinese provinces and counties, has introduced social networking to strengthen the relationship between the teacher and the learner networks, and demonstrated the effectiveness of this experiment studies. Runguaa Huang points out, that experience «broke teaching routine, and became more creative student. He adds Seaman Cheung, in his recent book, entitled (hybrid education): «These networks have a great ability to communicate information to the mind of the recipient with ease and spontaneity, and I call on all our educational institutions to expand in this experiment, and careful consideration in the psychological and practical aspects, this would remove the sense of alienation that may haunt some of the students inside the classroom, and make it possible to overcome the barriers that sometimes separate the students and educational institutions to which they belong, although we at the same time does not call for the absolute overshoot the official nature of the conduct of lessons, but done in a way balanced to ensure stability within educational institutions «Confirms an expert on modern education Philip Chang: The gradient in the use of social networks, within the Chinese educational scheme, can bring more benefit to the students, where the principle provides a stimulus and carrots, and ensures their access to the greatest enthusiasm, especially when it comes to lessons some complex may be deemed, for example, the lessons of learning foreign languages, which are mainly dependent on the openness and dialogue, and harmony within the school environment».

8.2 Case study two: *Philippine experience*

Philippines still many assumptions of parents that social media have negative effects on the academic performance of the student. On the other hand, there are parents who allow their children to be exposed to social media because they believe it will help their children be more alert and intelligent and excel in school. So far, schools and institutions experiencing many of the technological developments and changes in technologies to improve their teaching to meet the growing demands of the students methods. There are schools that benefit from the use of social media as a tool to assist in education. Through the study of the relationship between social media in relation to the academic performance of the students of Bachelor of Science in Information Technology at the University of Butterfish center-Malolos. Between one and thirty-eight (138) participants drawn using a stratified random sample there the first 45 years, 0.37 second and third year, and the fourth 19-year-old. There are 102 males and 36 females participated in this study. Reached seventy-one (71) or 51.4% of respondents said that less than satisfactory average grade while sixty-seven (67) or 48.6% of the students have reached a satisfactory academic performance than the average score successfully. Sixty-one (61) or a 44.3% fall in the occasional user of the social media framework while Seventy-seven (77) or a 55.7% fall in the frequent user of social media framework. This represents that there are more frequent than social media user among the participants in this study took part. Students associated with the initial reading and the average mid-term and on the use of social media by using the Pearson (r) correlation and evaluation that there is a statistically significant relationship between the moderate social media and academic performance of students from the University of BSIT Butterfish Center—Malolos, Bulacan.

8.3 Case study three: *American experience*

A newly applied in many schools and public and private institutions, and practiced widely by teachers and students, and was a recent study by a group of researchers at the University of Minnesota, have led to that 77% of the general students entering the networks with the aim of learning and the development of skills and openness to new perspectives, and that according to data collected over the last six months for students between the ages of 16:18 years, show «that students who use social networking sites, their skills and their creativity has evolved well», suggests Christine Jericho researcher the study, that the inclusion of the curriculum in social networks, helped to make the school more relevant

and meaningful for students, and became teachers are able to increase students' involvement in education, and raise the technological efficiency, and enhance the spirit of cooperation in the classroom, and build skills better communication», he adds: «the thinking not only in the integration of your technology, but in the creation of the most pressing tasks, and will develop critical thinking and problem solving, the ability to universal participation among students».

8.4 Case study four: *United Arab Emirates experience*

Having a proven track record in effective education, the Council of Abu Dhabi Education, is heading to the expansion of the use of social networking in the educational process networks, according to Director General of the Council, these networks have become an integral part of student learning, and to promote their association with local, regional and world ocean as a whole, and made them aware of all that the world is witnessing the latest developments, technical and scientific, cultural, so there is a strong trend to equip all schools, technical and educational paper scientific, entitled (the use of social networking sites in the classroom) was discussed in the Council, stated: «the need to support the transition to interactive education, especially in the classroom, and that many educators have become benefit from the media to achieve their educational goals, as it enables social networking activities to focus on the research and data collection and communication with the experts, and that You can use blogs to stimulate discussions and constructive dialogues and mutual cooperation in the electronic knowledge sites, and generally provide all of these social media easy access to support and exchange of experiences and professional development, and best practices within the professional and scientific society». The council has launched at the beginning of this year (2012 m), the project «electronic grade», in six schools in the Emirate of Abu Dhabi, the third grade students and the fourth basic education include the first episode, by two schools in each school district for a period of one year, and will be linked to every school from the six schools network «video-conference» electronic boards touch to encourage teachers and students to exchange knowledge and information on the local and global levels, to be circular in stages in the rest of the public schools in the emirate».

8.5 Case study five: *Oman experience*

Oman applied social networking to achieve the required educational implications of social networking sites on the undergraduate students in

the College of Applied Sciences (CAS), Nizwa, Oman. Blogs, wikis, tweets, RSS feeds, discussion boards, podcasts are educational contract in a huge network. Study charted uses this web2.0 applications and their impact on the linguistic and social behaviors for young students. Demographic segmentation builds a framework for assessing social tools and technologies, e-learning popular among the educated. The results of empirical evidence exploring the classroom and social programs such as forms that build young people knowledge societies. It examines variables that examine the effectiveness of these social tools in the exchange of knowledge and public awareness of the communities, and to try to achieve the following objectives:

- To discover how to use the students of the academy, Nizwa social networking sites?
- To determine how social networking sites can be used as a platform for learning academic education for students?
- Study of uses and gratifications drawn from the media and social impact of learning Students, on the whole, was made in response to more scoring for the use of the internet to write class assignments and the inclusion of student focus groups with college related social networks. Observations derived from experimental data show that students have begun using social networks for academic purposes. While some students SNS is seen as a distraction, and was reluctant to share their feelings, a high proportion of respondents found it a way to search for information, and to join the educational networks and search for job opportunities. The use of SNS gave them a sense of belonging to the academic community, as well as their friends online are mostly those who met in college. Two step flow interactions, student to student and teacher to student preferred academic learning through social networks. While the implementation of the social networking application in teaching after CAS Nizwa applied this method only once during the campaign in the YES WE CARE.

Beneficial results of this campaign, and uses and gratifications highlighted in this study, social networks and show a significant impact in the academic performance of students. However, the use of these networks to be disciplined because they can lead to distraction from education. The study also concludes that a significant portion of the students to take advantage on the importance of the human face to face classroom instruction, on social networks should be used for educational/ tutoring be able to apply these principles in a virtual classroom.

8.6 Case study six: Malaysia experience

Study in the Philippines conducted in order to shed light on the social media capabilities in academic circles through collaborative learning and improve academic performance for students. The researchers in this study using social media to improve the academic performance of students through collaborative learning among students and they are as follows with the interaction with peers, interact with the teacher and participation. In the results we have obtained, it may be concluded that social media facilitates the academic experience with the vast majority of participants, but you need to monitor and manage their time. Otherwise it will affect the use of social media negatively on the academic performance of students.

In line with the results of this study to understand the academic performance of students using social media in higher education, and after it was discovered: to gain in satisfaction of social media because they encourage and facilitate the use of student social media cooperative learning, and promote education and experience with the students. In terms of interaction with their peers on social media and got the highest ratio when it comes to the academic performance of students at the University because it affects to be simple for the student to go over the questions with other students through social media it is easy to communicate with peers and interact with them because It is the same age, class and education.

In terms of academic performance of students with participating got a typical ratio when it comes to the academic performance of students at the University because it provides the alliance and the exchange of knowledge in the classroom, the library using social media at any time. Cooperative learning with the use of social media and got a percentage of the average when it comes to the academic performance of students at the university. Because it helps to make the students feel confident enough to offer social media through a collaborative between colleagues, teachers and participating in class. Finally, in terms of academic performance of students with interaction with teachers from using social media and got the lowest percentage is not allowed in some cases to communicate with teachers or students are shy, but is good because those teachers provides a more understanding, and academic achievement in education.

CONCLUSION

The social media have infiltrated the 21st century generations of Internet users, making it a very active means of communications, particularly

among students of higher institutions of education. Consequently, academic activities in institutions and faculties are increasingly carried out through the social networks, such as Facebook, twitter and LinkedIn. These are essentially used in order to connect with current and prospective students and also to deliver instructional content. The results attained suggest that research model provided a powerful explanatory energy of social media academic satisfaction for among students. In the results acquired, it may be concluded that social media facilitates the academic experience and collaborative learning with the majority of the participants. The participants may also access material in addition to supplying sufficient content associated with their demands with peers and instructors. Fully engagement in classes towards improving academic performance using social media through collaborative learning in relation to interactive with peers, interactive with teachers, engagement, has resulted in persuasion to ease of use and perceived usefulness is hereby revealed.

REFERENCES

- Alexander, Bryan, (2009). Social Networking in Higher Education, accessed from <http://net.educause.edu/ir/library/pdf/PUB7202s.pdf>, 20 September 2011.
- Hamid, S., Chang, S. & Kurnia, S (2009). Identifying the use of online social networking in higher education. In Same places, different spaces. Proceedings ascilite Auckland 2009, accessed from <http://www.ascilite.org.au/conferences/auckland09/procs/hamid-poster.pdf>, 21 September 2011.
- Gruber, Abe, (2009). Social Media in Undergraduate University Admissions. Thesis of M.B.A. at Hawaii Pacific University, Honolulu. Accessed from http://www.bloomfield.edu/socialmediathesis/AbeGruber_SocialMediaThesis.pdf, 20 September 2011.
- Hermida, Alfred, (2011). Social media is inherently a system of peer evaluation and is changing the way scholars disseminate their research, raising questions about the way we evaluate academic authority, accessed from <http://blogs.lse.ac.uk/impactofsocialsciences/2011/06/27/social-media-is-inherently-a-system-of-peer-evaluation-and-changing-the-way-scholars-disseminate-their-research-raising-questions-about-the-way-we-evaluate-academic-authority/>, 22 September 2011.
- Oradini, F., & Saunders, G. (2008). The Use of Social Networking By Students and Staff In Higher Education. Paper presented at the iLearning Forum, 2008, Paris. Accessed from: http://www.eifel.org/publications/proceedings/ilf08/contributions/improving-quality-of-learning-with-technologies/Oradini_Saunders.pdf, 23 September 2011.
- Qualman, Eric, (2009). Statistic Show Social Media is Bigger Than You Think, accessed from: <http://www.socialnomics.net/2009/08/11/statistics-show-social-media-is-bigger-than-you-think/> 22 September 2011.
- Reuben, Rachel. The Use of Social Media in Higher Education for Marketing and Communications: A Guide for Professionals in Higher Education”, accessed from <http://www.ciff.net/DocumentoSeminararioII.pdf>, 21 September 2011. http://www.socialnetworkingwatch.com/all_social_networking_statistics/, accessed at 20 September 2011, <http://id.berita.yahoo.com/indonesia-urutan-ke-2-terbesar-pengguna-facebook-025416888.html>, accessed at 22 September 2011, “Social Media”, accessed from http://en.wikipedia.org/wiki/Social_media, 20 September 2011.
- Waleed Mughahed Al-Rahmi & Mohd Shahizan Othman; The Impact of Social Media use on Academic Performance among university students: A Pilot Study, *Journal of Information Systems Research and Innovation*; <http://seminar.utmspace.edu.my/jisri/>
- Brown, S. (2010). From VLEs to learning webs: the implications of Web 2.0 for learning and teaching. *Interactive Learning Environments*, 18(1), pp. 1–10.
- Schroeder, A., Minocha, S., & Schneider, C. (2010). Social Software in Higher Education: The Diversity of Applications and Their Contributions to Students’ Learning Experiences. *Communications of the Association for Information Systems*, 26, Article 25(1), pp. 547–564.
- Ferdig, R.E. (2007). Editorial: Examining Social Software in Teacher Education. *Journal of Technology and Teacher Education*, 15(1), 5.
- McLoughlin, C., & Lee, J.W.L. (2008). The Three P’s of Pedagogy for the Networked Society: Personalization, Participation, and Productivity. *International Journal of Teaching and Learning in Higher Education* 20(1), pp. 10–27.
- Wheeler, S., Yeomans, P., & Wheeler, D. (2008). The good, the bad and the wiki: Evaluating student-generated content for collaborative learning. *British Journal of Educational Technology*, 39(6), pp. 987–995.
- Kaitlin, C. (2010) Social Media Changing Social Interactions. *Student Journal of Media Literacy Education*, Issue 1, Vol. 1. pp. 1–11.
- Asur, S. & Huberman, B.A. (2010) Predicting the Future with Social Media. *Social Computing Lab: HP Labs*, Palo Alto, California. pp. 1–8.
- M. Owusu-Acheaw & Agatha Gifty Larson; “Use of Social Media and its Impact on Academic Performance of Tertiary Institution Students: A Study of Students of Koforidua Polytechnic, Ghana”, *Journal of Education and Practice*, Vol.6, No.6, 2015.
- Lenhart, A. & Madden, M. (2007). *Social Networking & Teens*. Retrieved from <http://www.pewinternet.org/Reports/2007/Teens-Privacy-and-Online-Social-networks.aspx>.
- Boyd, D.M. & Ellison, N.B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13(1), pp. 210–230.
- Raj Kumari Kalra & Preeti Manani; “Effect of Social Networking on Academic Achievement Among Introverts and Extroverts”, *Asian Journal of Social Sciences Humanities*, Vol2, No.3, August 2013 Vol.
- Saba Mehmood & Tarang Taswir; “The Effects of Social Networking Sites on the Academic Performance of Students in College of Applied Sciences, Nizwa,

- Oman.”, *International Journal of Arts and Commerce* Vol. 2 No. 1 January 2013.
21. Josan D. Tamayo, Giselle Sacha G. dela Cruz; “The Relationship of Social Media with the Academic Performance of Bachelor of Science in IT Students of Centro Escolar University Malolos”; *International Journal of Scientific and Research Publications*, Volume 4, Issue 5, May 2014.
 22. Nonaka, I., Takeuchi, H.; “The knowledge-creating company. How Japanese companies create the dynamics of innovation, Oxford University Press, Oxford.; 1995.
 23. Akubugwo, jeoma G1 and Maria Burke; “Effect of social media on post graduate students during academic lectures and library session. A case study of Salford University Manchester, United Kingdom”, *IOSR Journal of Research & Method in Education (IOSR-JRME)*, Volume 3, Issue 6 (Nov. –Dec. 2013), pp. 44–50.

Promote the process of higher education based on social media as a creative ICT tools

Ali Almajhaddi & Saad Almutairi

King Abdul-Aziz University, Jeddah, Saudi Arabia

ABSTRACT: Nowadays the use of ICT technologies, and more specifically computers, is increasingly fast in all activities. Connected to the increase use of new technologies is also the use of social media. New forms of connection spread among the internet, occupying an empty space of the web and also significant time of youngsters, becoming a daily routine for all of them. The social media communication triggered a considerable advance in social behavior, and originated several social media. Examples are blogs, wikis, online social networking and virtual worlds, some of these increased rapidly and remain nowadays, other succumbed to the natural evolution. Following [1], these social media are characterized as being social structures composed of persons or organizations, connected by one or more types of relationships that share common goals and values. The educators have always been early adopters in using new technology within their field. The educational technology such as Content Management Systems (CMS), Blackboard, and WebCT also used to help students perform better, as well as increase their productivity within the classroom. The aforementioned technologies enabled educators and students to better manage the learning process and share knowledge. This paper shed the light on the importance of social media in enhancing the education process either preliminary or higher levels. Also, our article presents some of the important experiments that show the enhancement process for various types of learning process through providing the educators new technologies to be used like: file uploading, discussion boards, and chat room services to streamline and enhance the education processes. Simple tasks and services such as giving out assignments electronically over the past few decades the Education began to look for new technologies as a tool to be used in order to improve the teaching process. We are now faced with another major challenge, the evolution and exponential growth of social media. The following articles try to answer the following questions:

- Social Medias may be one more step on the staircase of using new technologies in teaching and learning process?
- Do these can be useful tools to achieve the proposed objectives, as well the creation of rich moments of learning?

Keywords: content management systems, web based, social network system, webct, 5 cs, social network system and M-Learning

1 INTRODUCTION

Web is a democratic, open and DIY (Do-It-Yourself) communication media, as mentioned by [2]. The use of online social media networking activities in higher education varies from blogging, Wikis, and social networking sites. Table 1 shows summarize of online social media usage in higher education. [3]. Social Network (SN) media had shown their growth rapidly in higher education where the most users are at young age and “digital native”. According to this trend, some universities had adopted this opportunity to support their academic activity, official or unofficial activity, by students and lecturers. In the other site, some of them still worry about negative impact of this trend and choose to apply restricted access on SN

media. This research has an objective to explore the usage of SN media in higher education and try to uncover the benefit of SN usage to support academic activity.

The characteristics of social media can be summarized by the 5 C’s communication, collaboration, community, creativity, and convergence. With regards to communication; social media technologies are concerned with communication between and among human beings. This communication may be uni-bi-or multi-directional, collaborative, networked, or viral. Blogs may be viewed as an alternative or complement to publishing, but they may also be alternatives to personal web pages. Bloggers don’t only engage in one-way posting. Many cite each other’s work a great deal and post comments and ripostes on each other’s blogs, and

Table 1. Matrix of online social networking and social technology.

Social Technologies	Online Social Networking				Example
	Content Generating	Sharing	Interacting	Collaboratively Socialising	
Blogs	✓		✓		Blogger, wordpress
Wikis	✓	✓		✓	Wikipedia
Photo Sharing	✓				Flickr
Video Sharing	✓		✓		You Tube
Podcasting	✓	✓			
Social bookmarking	✓	✓			
Online discussion board	✓			✓	
Instant messaging	✓			✓	Yahoo Messenger, Google Talk
Social network sites	✓	✓	✓	✓	Facebook, Twitter

this results in “conversational blogging” [4]. Social networking sites like Facebook and Twitter enable communication among groups of people, large and small. Regarding collaboration; there is new media technologies enable collaboration over the Internet. Blogs in general have limited collaboration, although a single blog may be shared among a group of bloggers and sometimes a blog may be used for group work. Wikis are today’s collaboration tool. These are also scalable, in that private wikis for small groups of people work just as well as the largest collaborative product we have ever seen—Wikipedia. Some authors have examined how wikis are used [5], both at work and in other arenas of life. Social media also encourages collaboration with virtual conferencing on.

The use of SNS has been found to have positive effects with the development of personal identity, self-esteem and social capital. Social networking sites are being used as a place to build personal identity, self-esteem and social capital, however along with the benefits of these sites come new concerns, issues and anomalies. SNS have become staging grounds to test and implement personal identity among youth who desire a safe zone where they can experiment free from the scrutiny of authority figures. An increased level of self-esteem, especially among users with poor self-esteem, is also a trend that is stemming from increased use of Facebook and other sites. In addition, the development of social capital seems to be more heavily reliant upon small network clusters, which are just as prevalent offline as online, revealing that SNS

are only of small value for building social capital. The social media technologies often referred to as Web 2.0 encompass a wide variety of web-related communication technologies such as blogs, wikis, online social networking, virtual worlds and other social media forms. Much has been said about the unique character of the social media technologies, the features that unite these seemingly disparate technologies under a single umbrella [6].

The objective of this article is to uncover some fact about SN usage in higher education by identify the dominant usages on basic functionality and academic specific activity. The result is expected can answer such questions as:

1. Are there positive benefits of social media usage in higher education?
2. What of dominant functionality of social media usage in higher education?
3. What kind of policy and support needed to optimize the benefits of social media usage?

2 BACKGROUND AND PRELIMINARIES

The last few years have been really significant in terms of technological evolution of our society. The teaching-learning process was, over time, mutating, evolving, adapting, and always considering the changes occurred in contemporary society. Teaching, generally, reflects the society, but, in a particular way, the teaching process presents the individuals modus operandi that live in this society and the available resources. Teaching in

any room whatever the content, Mathematics, Languages or even Sciences, it is not a process that should always follow the same guidelines or be a part of structural or contents changes, given the important changes observed in recent years in this regard. The emergence of computers and other more recent technological tools, as smartphones and tablets, pave the way for a world so vast and diverse, presenting fruitful possibilities of acquiring information and other situations that belong to a dimension more futile, regarding the veracity of the subject observed. One of the major implications of the rapid development of these instruments was the enlargement of the normal context of what is a classroom. Defining classroom could be more difficult than before, because this is not the only place educator must have as exclusive to teaching-learning process: nowadays, it should be given the chance to the students to have access also at home, or anywhere, to direct support by the educator and even by their own colleagues. Aware of this trend, educators were adapting. Initially, somewhat slow, just using the computer for smaller jobs; more recently, the exploration of internet connections and, as a consequence, the use of resources obtained through this means. As such, in this vast ocean of knowledge accumulated, it becomes essential to understand the possible choices for a better understanding, so that subsequently we can better understand what to do and how to do it [7].

Introducing the specific area of mathematics, there are many available software, author or free ware, however, the vast majority have closed structures. Other possibilities such as dynamic geometry software, intend primarily for the specific area of geometry, dimensions, presenting itself, therefore, as a tool with a single objective, i.e., reductive. Nowadays, saying that it is possible to teach differently in a math class is complicated and does not truly defines “differently”, because there are few changes from a normal process of teaching. Besides these factors, there is another, probably the most important: the interaction of the student with this kind of technology and software. Students are not always motivated to work with these tools, making this a key aspect for the presentation of new technologies in the context of the teaching-learning process. The motivation, essential, whether in a traditional classroom exposure of content, either in a class that makes use of the technological elements, it is imperative factor for success [8].

3 LITERATURE REVIEW

There is pros and cons of Social Network (SN) usage in higher education environment. Although some of Web 2.0 users are youngster, but to adopt

the social technologies in class room needs careful plan for some reason, for examples not all digital natives are keen to have such technologies for various reasons: diversity of experiences, familiarity, attitudes and expectation of the students towards online technologies [9].

Despite of the pros and cons, some of social network media has exceed their functionality, from non Formal media into formal and official media, both for public information and communication. Facebook and Twitter for example. Facebook itself, after 3 years release (2007), this site introduced the fan page concept which is successfully attracting many users to create their fan page officially. The owner of fan pages varies from small or individual business into big company such as HP, Fujitsu and so on. Many universities jumped at the opportunity to create an official Facebook, and also some of sub organization in the university, for example library, students association, student’s forum, etc. Fan pages can creates viral marketing effect, because it can catch the interest from the friends of people who already become fan [10]. Twitter also has successful reach the popularity relatively faster than Instant Messaging such as Yahoo Messenger (YM!) or Google Talk. Having Twitter account can become one of “official” identity in virtual worlds, beside email address as you can see in formal poster publication in public area. Many students and lecturer creates viral networking using twitter and follow each other’s. It also helps the user to spread information and express their opinion directly to the others.

The usage of SN media in higher education can vary from marketing media, information media, communication media, feedback, complain, announcement, sharing, task assignment and examination. The intensity of SN media usage depends of some factor such as:

- The background and behavior of user.
- University policy on internet access.
- The behavior of university communication.
- The role and rule of SN in daily communication.
- The attitude of user.

4 SOCIAL MEDIA FOR ACADEMIC PURPOSE

When studying relationship between social media and higher education apparently, social media is being used increasingly by university students. It is promoting virtual communities and Virtual Learning Environments (VLEs) for expanding [11] distributed learning among users. The students interact in their virtual communities freely with members

of the community. They can share information and study experiences, research projects and job opportunities with each other. Various factors contribute towards the use of social media for educational purposes. Armstrong & Franklin (2008) compiled a comprehensive report 2008. The report indicated that the students used social media in different manners to enhancing and strengthening their learning, through reflection and collaborative activities in virtual environments. However, they depended upon infrastructure including and the skill of using social media [12].

The continued growth of social media presents a set of clear challenges to the future nature of higher education provision and practice. Yet as with many previous new technologies, academic discussion and debate remains largely speculative rather than well-informed and certain. Of course, there is an emerging literature of small-scale, “empirical” studies that confidently reports all manner of specific learning gains and benefits from social media. We have been recently told, for example, about the positive effect of Twitter use on college student engagement and grades [13], and the ability of social networking sites to engender “favorable feelings regarding learning experiences” [14]. Yet, rather than being a wholly good (or wholly bad) thing for higher education, social media are perhaps best understood in more ambiguous terms. This is especially the case when one considers the complex and often compromised realities of students’ actual uses of social media within educational contexts and in their wider everyday lives.

Now a days, the ever increasing use of social media at higher education level seems to be transforming the prediction of [15] that “Universities will lose their privileged role as a primary producer of knowledge, and gatekeeper to it, as knowledge becomes more widely accessible through other sources and is produced by more people in more ways into reality The usage of social media by university students is an interesting area of research for educationists and social scientists. [16], [17] were of the view that the available literature contains beneficial designs and styles of using it at university level. It describes the creation of contents and less focus on how to share, interact, and collaborate and socialize by its use. There seem different reasons to justify the usage of social media in higher education. It usage was affirmed by upholding the stance that it is used to enhance study experiences of students by provision of e-support services to them [18]. It is used to facilitate communication among and between students in virtual communities. Amongst others, the Facebook appears to be the most favourite was suggested as a means of communication for interacting with students [19].

There is also growing evidence that social media use is not the equitable and democratic activity that it is often portrayed to be. Even when able to access the technology, the types of social media tools that an individual uses, the ways in which they are used and the outcomes that accrue are all compromised by a set of second-order digital divides. For instance, recent studies suggest that students’ preferences for particular social media applications over others follow class-based patterns of taste and distinction. In terms of social networking, for example, [20] reports that US college students’ preferences for an application such as Facebook as opposed to My Space appear to be patterned consistently along lines of social class and educational background. Clear socio-economic differences also exist in individuals’ predilection to produce (rather than consume) online content, be it posting to blogs, sharing resources or creating profiles [21]. Other US college studies report that social media environments are no more socially integrated than offline contexts. For example, race has been found to remain the overriding predictor of whether college students are Facebook “friends” or not. Similarly, social media do not necessarily overcome issues of offline disabilities but instead often exacerbate the boundaries of disability [22]. All in all, it is unhelpfully idealistic to imagine social media as providing a level playing field for all.

The present time is regarded to be the information age providing open access to all. The younger generation called Net-Generation appears to be much inclined towards having information by using modern technologies. Educational usage of social media seems useful for all levels of education but university students are much crazy to use it [23], [24]. Social media can be said to be the communication facilitator and students wish their institutions to use social networking sites for strengthening classroom [25] instruction. In this [26], [27] stated that they lead to use social media to enhance educational access and interaction. Moreover, social networking can fill the learning gap informally between “digital native” students and “digital immigrant” faculty.

Many of the current discussions and debates over social media are also (deliberately) unclear as to what aspects of social media use actually relate to education, learning and knowledge. One study of United Kingdom students’ use of Facebook suggested that the vast majority (around 95%) of students’ interactions were completely unrelated to their university studies. Thus while social media may well have the potential to support communal learning and knowledge generation, this is by no means guaranteed. In this sense, [28] make the useful distinction between living technologies (i.e. technologies that students choose for their everyday

social lives and for leisure purposes) and learning technologies (i.e. technologies that students use primarily for study purposes). As this distinction suggests, while there may be some overlap between the two, we should not mistakenly presume all of the everyday life aspects of social media use to be of educational significance.

The main role of social networks is to coagulate virtual learning communities within the scope of discussions on topics such as scientific subjects, virtual experiments or other various themes as exams preparation. A major impact in promoting these activities focused around the concept of “social networking” is provided nowadays by the development of client applications for mobile devices which enhances their accessibility. Moreover, the development of technologies needed to implement e learning software on the latest generation telephone devices and other mobile devices has led to the shaping of new concepts such as mobile-learning or, in learning foreign languages, Mobile Assisted Language Learning (MALL). In the area of social networking, recent surveys highlighted that 30% of Facebook users and 37% of Twitter are using the networks from mobile devices.

Finally, it may also be a mistake to presume that students are necessarily enthused and motivated by the use of social media. For example, [29] study of US College students found that a significant proportion made little use of social media applications, with students as likely to be dabblers or outright dissenters as they were to be social media omnivores. In this sense, it is unwise to assume that the interest, motivation or affinity of all students will be enhanced by the inclusion of social media technologies in any educational context. Indeed, a number of commentators warn against attempts to motivate and engage students simply through the introduction of consciously trendy forms of social media technology use into educational processes and practices. As [30] conclude with regard to the (mis)application of social media tools in the workplace, young people’s ‘appetite for authenticity means that they are resistant to ill-considered attempts by older generations to “speak their lingo”’.

5 WELL KNOWN SOCIAL NETWORKING SITES

Formerly designed on communication purposes and for improving information exchange among small groups of users, social networking sites have become quickly very popular, and the number of users from a wide geographical area joined the groups and became regular clients. In general, the social networks sites provide users with a private

virtual space where each one could build his own public profile and manage a list of links to other users’ profile.

5.1 Facebook (*Facebook.com*)

Facebook may be the face of online social networks. Developed in 2004 by then Harvard undergraduate Mark Zuckerberg, it is the “dominant” social networking site. Among the many studies reporting statistics related to Facebook adoption and usage, Ellison, Steinfield, and [31], found that 94% of their college students were users of Facebook spending an average of 10–30 min on the site and having 150–200 friends. More recently and in a larger study, 90% of undergraduate college students were reported to have Facebook accounts. Reaching the one billion user mark during the first days of October, 2012 (Facebook.com), an interesting usage trend has recently emerged. College-age users (Facebook’s most mature market) reportedly spent 25% less time on the site in August of 2012, a declining trend predicted to continue.

Despite the worldwide spread of Facebook users, there are still countries in the Middle East (or even China) where Facebook is banned or limited. In terms of educational impact on higher education institutions, at the moment there are several institutions registered on Facebook, but also students, parents and many groups specially created for finding school or university colleagues.

Actually, on Facebook we find all forms of interaction between educational services providers, direct beneficiaries of education services, and why not, parents of students as stakeholders. These interactions can take several educational approaches for Facebook users: (a) Learning for using Facebook. (b) Using Facebook for learning. (a) Learning for using Facebook could be considered a strange approach, but this is a concept which emerges from the users’ incontestable interest in own information security and privacy in order to answer to questions such “What could happen when a student makes public his/her information on Facebook?” On the facebookforparents.org website tips and good practices are available for parents, in order to keep safe the children while they are surfing on Facebook pages. Things have gone further and there are software applications designed for data security which provide the option to deny access to Facebook to specific users of a given computer. However, a proper understanding of social networking concept and a proper evaluation of knowledge spreading potential could be an important step for decision makers in network security for many institutions. (b) Regarding from the Using Facebook for learning point of view, teachers seem to be less convinced than students

to use Facebook. The teachers' reluctance on using Facebook to communicate with students is not probably resulting from their conviction that using Facebook would not produce beneficial effects on learning, but from their concerns about security of information conveyed in social networking and high exposure on the Internet for teachers' privacy. There are many possible uses of Facebook in education, some authors [32] stating about 100 ways to use Facebook in the classroom, in order to provide value to the educational process. The main features which recommend Facebook as a valuable tool which could be used in education are:

- Teachers can create custom list of students and manage groups of students on custom topics related to courses;
- Exchanging information through links, photos or multimedia content related to specific subjects;
- Creating surveys and quantifying the feedback.
- Using the on line chat for direct communication between students and teachers.
- Publishing news on tests, exams or face to face meetings.
- Integrating Facebook with other collaborative services provided by other application (like Google docs).
- Using Facebook as a complement for an eLearning platform.

5.2 *Twitter (Twitter.com)*

Twitter is a social networking site that is often termed a microblogging service. In contrast to Facebook or My Space, Twitter limits posts or updates to 160 characters. Some have suggested that Twitter makes for a faster mode of communication because of the relatively short post lengths. The average blogger may update every few days whereas the average microblogger will update several times a day. There is no doubt that Twitter has become of great importance in the professional life of teachers, both in terms of its use as an educational tool or as a means to connect teachers with each other, to keep up with them for the trends of modern education and the growing association with the technical world of many examples of this, including the personal learning network PLN which is considered a model embodying this provided the opportunity given to teachers in order to access another global levels of learning technologies in order to renew the education away from the typical and classic, [33] and many assert that Twitter is the best form of platforms and networks, it also provides cooperation between experts across the world, and gives the opportunity to see the codes of teachers and use them in their classrooms,

as resource sharing easily and in less time, and is Twitter as a meeting room HYPOTHETICAL to communicate with co-workers and get the latest news and all that is new in the field of education.

You can determine the uses of Twitter in Higher Education as follows:

- * Twitter as pallets ads: you can use Twitter to place ads for your students to your followers example: the news of the postponement of the test date or change the date of a lecture or a request for a new search mode.
 - * Twitter as a tool Review: Create marking or Hachtaq as the article or unit (for example: # revision _part_ one), and publish it to the students in which to discuss or review the content of this unit.
 - * Twitter as a tool to support office hours: you create an account on Twitter can help your students to communicate with you to inquire about a particular topic or discussion around a point.
 - * Twitter as a tool to coordinate and follow-up projects: instead of sending e-mails to students or wait until the lecture or the next portion of the discussion or keep track of the students' work on a particular project, students can work on Twitter and create a marking or Hachtaq for their project and it will be followed up their activity and keep track of developments on their projects.
 - * Twitter as a tool to break down barriers: shame and dread spread among some students have student ashamed of the question or direct the discussion in front of everyone and Twitter could help them break this barrier.
 - * Twitter as a tool to communicate with parents: Parents may use Twitter to follow their children's teachers and to stay abreast of the latest activities of their children and their experiments and projects.
 - * Twitter as A Digital Masters: Twitter can be made a tool for discussion between the teachers and the teachers and the participation of diverse and useful resources.
 - * Twitter calendar tool: Try using Twitter with your students in the evaluation of their information about the last lesson may be to allocate an hour a day for it, do not forget to respond to them directly feedback are the basics of the evaluation process.
- Sudden activity on Twitter: Try asking surprise questions on Twitter, and give additional degrees of answer faster.
- * Twitter as a tool to collect and share Sources: Ask students to share resources or additional information on the topic of your lesson and share in it.

- * Twitter as a tool to communicate with experts: Use Twitter to search for instructional and pedagogical experts and follow-up new and take advantage of their expertise to develop your skills.
- * Twitter as a tool for brainstorming: you can share ideas and information with your students at any time.

Use the site (twtpoll.com) To create a poll or vote and participation in Twitter to know the opinion of students in a particular subject.

- * Twitter as a tool to get to know others: Look with your students about another teacher of the same substance and try his participation and his student's information and discussions.
- * Exploit Twitter for students frequent discussion and interventions, the Marks in a classroom student much debate and provincial information may not have irrelevant, it has been used (Jenny Robinson) a teacher—Twitter with one of the students with autism for this purpose was receiving the discussions in Twitter and respond to useful tweets and leave what is not useful.

6 CURRENT USES OF SOCIAL MEDIA AT THE HIGHER EDUCATION

6.1 *Library uses*

A growing number of college libraries are tapping into Facebook and My Space. At Georgia Tech, the information services librarian reports using Facebook to network with mechanical engineering students [34]. "With the undergraduate enrollment for mechanical engineering around 1,700 students, I was surprised to discover that more than 1,300 of the users on Facebook. This presented an intriguing opportunity to directly market the library to more than 75% of my target audience" [34] felt that Facebook had help accomplish his goal of promoting the library as a subject liaison and helped meet the needs of his students.

However, librarians also report hurdles to be overcome when using Facebook. [35] Explain that, "Unfortunately, Facebook firmly outlines in its terms of use that a group or entity cannot register a user Profile, so it deleted them. Facebook managers encouraged librarians to replace the deleted institution user Profiles with new personal Profiles and to form Groups to promote library services to patrons (Facebook seems to have changed this rule since that time.) [35] also utilize Facebook in their library and argue that "You have to connect with your patrons before you can effectively promote your services to them. Many libraries across the United States are using My Space and

Facebook to reach students, announce library events, and answer research library-related questions. [36] explains this phenomenon, contending, "Many academic libraries have developed a presence in online courseware with links to library services targeted to online learners. Similarly, the Brooklyn College Library has provided a MySpace portal to its services ... that contains links to the catalog and databases, as well as, documentation on how to access library resources off-campus. In fact, in one study, Librarians wanted to determine which source students would use more to ask reference and research related questions: email, phone, instant message, Facebook, or in-person [36] Students in this study preferred asking their reference and research related questions using Facebook and email even more than face to face. However, the picture is not all positive. Based on a survey she did of 366 Valparaiso University freshmen to gauge their perspectives on libraries using Facebook and My Space as outreach tools, [37] recommended that librarians "proceed with caution" on this decision. She found that though most students seemed open to the idea of the library contacting them in this way, some (12%) reacted negatively because of the potential to infringe on their sense of personal privacy. A previous study by [37] also reported some ambivalence on the part of students to using Facebook or MySpace to connect with libraries. Some noted that email seemed to them a more appropriate avenue for this kind of communication.

6.2 *Faculty uses*

SNS uses are also beginning to emerge in college classrooms. The [Facebook] network is increasingly being used not only by students but also by [college] faculty. According to a Facebook spokesperson, approximately 297,000 Facebook members identify themselves as faculty or staff" (p. 3). This has potential for both benefits and consequences, [38] say that "Students may perceive a teacher's use of Facebook as an attempt to foster positive relationships with his or her students, which may have positive effects on important student outcomes ... (However), teachers may violate student expectations of proper behaviors and run the risk of harming their credibility if they utilize Facebook. Despite this potential consequence, teachers may enhance their credibility among students by signifying an understanding of contemporary student culture", and students about faculty use of Facebook. "The primary purpose of this study was to explore the impact of teacher self-disclosure on Facebook on student motivation, affective learning, and classroom climate. Findings suggest that higher teacher self-disclosure may lead students to higher levels and affective learning and lend to

a more comfortable classroom climate”). In this study, SNSs seem to work to improve classroom climate for the instructor.

6.3 *Administrative uses*

The most prevalent use of SNSs in the university community is creating profiles and groups to communicate events with users. Colleges are also using SNSs for university marketing campaigns. Facebook seems to be perceived as “... an excellent mechanism for communicating with our students because it allows us to go where they already are; it is an environment that students are already comfortable with” [39]. Southern Illinois University’s College of Business at Carbondale reports using Facebook to communicate and market school events as well as activities to students and alumni. The university department reported 400 members on its Facebook group, allowing “... members to receive school news and communicate easily with students, faculty, alumni, and others in the school community”.

One university official explained the perceived benefit. “The group offers the school an easy, no-cost way to post school announcements, recruit for student organizations, and upload photos. In the future, the school may use Facebook to survey students on different topics. Facebook also makes faculty seem more approachable and opens up new avenues of communication” [40] Facebook, as well as other SNSs, have been used to open the communication lines between students and universities by informing them of college events and other collegiate activities.

7 BENEFITS, DRAWBACKS AND BEST PRACTICE OF SOCIAL MEDIA IN EDUCATION

You cannot look at the use of social networking sites on education as a positive or negative overall effect, but neutral as an influence. The way we use are determined by this effect, whether positive or negative, impact on our relationship. That means like as many of the common issues that can be utilized in health beyond grades, as can indulge in error beyond, the issue is not passive in the act and not the lesson of the user of the user, how many people took them networking from ignorance to knowledge]. That’s a lot of information bombarding students. Trying to keep up with it all can change the way the brain functions. Is this good or bad? Both. WCER researcher Mark Connolly acknowledges that these social media show value in educational settings as long as they are used prudently. Many have pointed to the educational

benefits of these media (also called Web 2.0). Social media tools and networking sites encourage students to engage with each other and to express and share their creativity.

7.1 *The drawbacks of social media using in education*

Seen that there may be some drawbacks to the use of social networking in education networks. For example, it may be used involves a violation of privacy, where a profile for each student there contains information about him and his whereabouts and activities and inclinations, have misused this information in the case of exposed persons other than trusted. Also, the use of the Internet to communicate without doubt reduces direct and personal confrontation between the teacher and his students, which can sometimes be important to create a strong and lasting relationship between them. One important flaw is that there may be scope for fraud if the use profile of a student by another student is the owner of the file [41].

Finally, the increase in the number of hours a student spends in front of a computer may lead to some psychological or social problems. In spite of these few disadvantages, the advantages seem much more, which leads us to believe that the role of social networking in education will continue, but will increase and become more widespread application and in the next few years. The education of social networking has locations flaws including: (uncertainty: sites, social networking is prone to penetrate at any time, and there is no guarantee that the account is familiar it is not due to impersonator recipe, ease of amendment: Very easy to author edit or delete Their publications on social networking sites, which poses a real problem in front of this kind of martyrdom, Add to that the lack of recognition: some of the research committees do not recognize already that kind of information, which is adapted from the social networking sites).

7.2 *Benefits of social media in education*

Advantages and benefits of the use of social networking in education: According to the evidence of modern psychology, the process of storing the human mind to the information, or vocabulary, is determined by its nature the psychological state of the recipient, and then the need for a recreational dimension during the lecture process, and this is made possible by networking sites social, where students will be more motivated, especially when learning languages, mathematics, social and material, [42] It is also active skills of learners, and provide opportunities for them, and encourage their

creative thinking patterns and different ways, and the pride of the positive role of the learner in the dialogue, and make it an active participant with others, and promote educational methods in a collaborative environment, and helps the learner to studying constructive by offering a variety of exercises and integrated and allows the teacher and the learner the possibility of exchanging books-and-follow new information in the specialty and provide educational games aimed at making use of opinion polls, where the teacher uses these surveys as an educational tool effective and increase communication with his students, and follow-up of new news and stand on the emerging current events political and economic, scientific, social, and instilling the ambition in the hearts of the learners by encouraging them to create and design new applications to networks serving the educational material, and disseminated among the educated to take advantage of them, where many of the students presented their applications process among themselves, such as the number of schools and international institutes students, who make up Groups on the site.

The education through social networks contribute to the transfer of education from the competition phase, the integration stage, by requiring all learners to participate in the dialogue and information gathering which makes teaching and learning more fun and energetic and experience the clock, and the participation of the challenge, where the teacher can engage his students in the implementation of projects related to promoting their institutions teaching, in order to measure their talents and enrich their abilities, and the extent of their self-confidence, and the introduction of new methods, are encouraged to put forward ideas, and promote a spirit of partnership and communication between learners, and to enable the teacher to put himself Office hours face hours, allows the students through which to communicate with him ask questions and receive answers, and facilitate the distribution of scientific subjects in the classroom in the process, as well as facilitate the evaluation process, and testing.

7.3 *Considerations for the educational use of social media*

Students must learn to distinguish the skill needed to locate information online from the ability to understand that information. Using social media to cultivate and demonstrate deep learning is possible, but that requires overcoming the persistence of distraction, the surfeit of irrelevant information, and the temptation to wander. Students can develop a capacity for practical reasoning when using social media. Educators and students should have multiple, purposeful discussions about social

media's pros and cons. Social media can enhance and impede student learning, and educators can use realistic case studies to help students identify trade-offs. For example, the use of social media in educational settings may incorrectly suggest that learning should be easy and quick [42]. If so, students should be shown the value of reinvesting the time and effort saved by technology into higher-order tasks that really matter to their learning, such as writing a complex argument, reading difficult texts, and debating ideas with others. Social technologies are here to stay. Connolly says that it is important to help students learn how to use social media in an instrumental way, learn how to think deliberately about their use, and consider the sorts of outcomes for which using social media are proper. In the real world, students will find themselves facing a difficult situation involving social media that rules alone cannot resolve. Connolly says the problem will require their best judgment—a kind of practical wisdom that cannot be taught, but instead is learned through practice accompanied by guidance and support. Knowing when, where, and with whom to use social media, Connolly concludes, may be the most important learning outcome of all [43].

8 CONCLUSION

Tools for education provided through social networking sites offer specific advantages especially for distance learning, using an affordable and popular environment. Currently, online social networks are used by heterogeneous groups with different ages which tend to integrate more and more facilities offered by these networks in their daily lives. There is no doubt that, with the unprecedented expansion of social networks, personal data security policies must to be improved and users are have to be better trained to protect themselves. In recent years several social networking users have been victims of hackers, spam, malware or phishing. Tools for education provided through social networking sites offer specific advantages especially for distance learning, using an affordable and popular environment. Currently, online social networks are used by heterogeneous groups with different ages which tend to integrate more and more facilities offered by these networks in their daily lives. There is no doubt that, with the unprecedented expansion of social networks, personal data security policies must to be improved and users are have to be better trained to protect themselves. In recent years several social networking users have been victims of hackers, spam, malware or phishing.

One of the most popular social networks in Romania is Facebook, which is gathering many

visitors, especially young people, from different backgrounds and continues to expand rapidly in all age groups. The rapid development of technology in the field of mobile devices is opening new opportunities for knowledge transfer and social networks are the first to benefit. Student are very receptive to the development of technologies for mobile devices and implementing e-learning software on the telephone devices or other mobile devices is already leading to the shaping of new concepts (as E-Learning or mobile-learning).

On the education level, educators who refuse to adapt and continue to insist that the only way to learn is via “chalk and talk” methods will find themselves hopelessly obsolete. Besides the changes that have taken place in libraries and in journalism, one only has to think of what is currently happening to the film industry where the traditional way of showing films in movie theaters is disappearing. Professors who wish to increase their value to their institutions must embrace technology and use all kinds of tools to impart knowledge.

REFERENCES

1. Armstrong, J. & Franklin, T. (2008). A review of current and developing international practice in the use of social networking (Web 2.0) in higher education. 2008. Retrieved 18 August 2011 from <http://www.franklin-consulting.co.uk>
2. Palen, L., Vieweg, S., Liu, S. & Hughes, A. L. (2009). Crisis in a networked world: Features of computer-mediated communication in the April 16, 2007, Virginia Tech event. *Social Science Computer Review*, 2009, 467–480.
3. Palen, L. (2008). Online Social media in crisis events. *EDUCAUSE Quarterly*, 2008, 31(3). Retrieved 2nd December 2011 from <http://net.educause.edu/ir/library/pdf/EQM08313.pdf>.
4. Smith, P. Smith, N., Sherman, K., Goodwin, I., Crothers, C., Billot, J. & et al. (2009). *The Internet in New Zealand 2009*. Auckland: Institute of Culture, Discourse and Communication, AUT University.
5. Hussain, I. (2005). A study of emerging technologies and their impact on teaching learning process; An unpublished PhD Dissertation, Allama Iqbal Open University Islamabad, 2005.
6. Armstrong, J. & Franklin, T. (2008). A review of current and developing international practice in the use of social networking (Web 2.0) in higher education. 2008. Retrieved 18 August 2011 from <http://www.franklin-consulting.co.uk>.
7. Hamid, S. Chang, S. & Kurnia, S. (2009). Identifying the use of online social networking in higher education. Same places, different spaces. *Proceedings Ascilite Auckland 2009*. Retrieved 18 August 2011 from <http://www.ascilite.org.au/conferences/auckland09/procs/hamid-poster.pdf>
8. Dabner, N. (2011). Design to support distance teacher education communities: A case study of a student– student e-mentoring initiative. *Proceedings of Society for Information Technology and Teacher Education International Conference 2011*. Nashville, TN: AACE 1-880094-84-3., 2011
9. Mack, D., Behler, A., Roberts, B., & Rimland, E. (2007). Reaching students with Facebook: Data and best practices. *Electronic Journal of Academic and Special Librarianship*, 2007, 8(2).
10. Davis, N. E., Dabner, N., Mackey, J., Morrow, D., Astall, C., Cowan, J., & et al. (2011). Converging offerings of teacher education in times of austerity: Transforming spaces, places and roles. *Proceedings of Society for Information Technology and Teacher Education International Conference 2011*. Nashville, TN: AACE 1-880094-84-3, 2011.
11. Roblyer, M. D., McDaniel, M., Webb, M., Herman, J. & Witty, J. V. (2010). Findings on Facebook in higher education: A comparison of college faculty and student uses and perceptions of social networking sites. *The Internet and Higher Education*, 2010, 13 (3), 134–140.
12. Madge, C., Meek, J., Wellens, J., & Hooley, T. (2009). Facebook, social integration and informal learning at university: “It is more for socializing and talking to friends about work than for actually doing work”. *Learning, Media and Technology*, 2009, 34(2), 141–155.
13. Bull, G., Thompson, A., Searson, M., Garofalo, J., Park, J., Young, J., & Lee, J. (2008). Connecting informal and formal learning experiences in the age of participatory media. *Contemporary Issues in Technology and Teacher Education*, 2008, 8(2). Retrieved from <http://www.citejournal.org/vol8/iss2/editorial/article1.cfm> [accessed 18 August 2011].
14. Olson, J., Clough, M., & Penning, K. (2009). Prospective elementary teachers gone wild? An analysis of Facebook self-portraits and expected dispositions of preservice elementary teachers. *Contemporary Issues in Technology and Teacher Education*, 2009, 9 (4), 443–475.
15. Grimmelmann, J. (2009). Saving Facebook. *Iowa Law Review*, 94, 1137–1206, 2009. Retrieved 2nd December 2011 from http://works.bepress.com/cgi/viewcontent.cgi?article=1019&context=james_grimmelmann.
16. Cain, J., Scott, D., & Akers, P. (2009). Pharmacy students' Facebook activity and opinions regarding accountability and e-professionalism. *American Journal of Pharmaceutical Education*, 2009, 73. Retrieved 4th September A. Hewitt, & A. Forte. *Crossing boundaries: Identity management and student/faculty relationships of 2011* from <http://www.ajpe.org/aj7306/aj7306104/aj7306104.pdf> Facebook (2006). CSCW06, November 4–8, Alberta, Canada, 2006.
17. Selwyn, N. (2009). Face working: Exploring students' education-related use of Facebook. *Learning, Media and Technology*, 2009, 34(2), 157–174.
18. Efimova, L. & de Moor, A. (2005). Beyond personal webpublishing: An exploratory study of conversational blogging practices, *Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05)*, p. 107a.
19. Tapscott, D. & Anthony D. Williams (2006). *Wikinomics: How mass collaboration changes everything*, New York: Portfolio Hardcover Publishers.

20. Johnson, M. J. & Wilcox, P. A. (2007). The world of connected things. *The Journal of Government Financial Management*, 56(4), Winter, 48–53.
21. Andrus, D. C. (2005). The wiki and the blog: Toward a complex adaptive intelligence community. *Studies in Intelligence*, September, 49 (3).
22. Hynes, D. (2003). Consumption convergence. *Irish Communications Review*, 9.
23. Papacharissi, Z. *A networked self*. London, Routledge, 2010.
24. Subrahmanyam, K. & Šmahel, D. *Digital youth*. Berlin, Springer, 2011.
25. Junco, R., Heiberger, G. & Loken, E. “The effect of Twitter on college student engagement and grades”, in *Journal of Computer Assisted Learning*, 27, 2, pp. 119–132, 2011.
26. Hung, H. & Yuen, S. “Educational use of social networking technology in higher education”, in *Teaching in Higher Education*, 15, 6, pp. 703–714, 2010.
27. Jones, S. & Fox, S. *Generations online in 2009*. Washington, DC, Pew Internet and American Life Project, 2009.
28. Helsper, E. & Eynon, R. “Digital natives: where is the evidence?”, in *British Educational Research Journal*, 36, 3, pp. 503–520, 2009.
29. Hargittai, E. & Hsieh, Y. “From dabblers to omnivores”, in Papacharissi, Z. (ed.). *A networked self*. London, Routledge, 2010.
30. Schradie, J. “The digital production gap”, paper presented to the American Sociological Research Association conference, San Francisco, August 2009.
31. Lewthwaite, S. “Student experiences of social networking and disability in higher education”, unpublished PhD thesis, University of Nottingham, 2011.
32. Hosein, A., Ramanau, R. & Jones, C. “Learning and living technologies”, in *Learning, Media and Technology*, 35, 4, pp. 403–418, 2010.
33. Shirky, C. *Here comes everybody*. London, Allen Lane, 2008.
34. Waycott, J., Bennett, S., Kennedy, G., Dalgarno, B. & Gray, K. “Digital divides?”, in *Computers and Education*, 54, 4, pp. 1202–1211, 2010.
35. Lee, M. & McLoughlin, C. *Web 2.0-based e-learning*. Hershey PA, Information Science Reference, 2010.
36. Nicholas, D., Gunter, B. & Rowlands, I. *The Google generation*. Oxford, Chandos, 2009.
37. Crook, C. “Theories of formal and informal learning in the world of web 2.0”, in Livingstone, S. (ed.). *Theorising the benefits of new technology for youth*. Oxford, Oxford University Press, 2008.
38. Leadbeater, C. *We-think*. London, Profile, 2008.
39. Hargittai, E. & Hsieh, Y. “From dabblers to omnivores”, in Papacharissi, Z. (ed.). *A networked self*. London, Routledge, 2010.
40. Tapscott, D. & Williams, A. *Wikinomics*. New York, Atlantic, 2007.
41. Boyd, M. D. & Ellison, N. B. (2007). Social network sites: definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230.
42. Negrilă, S. (2010) “Cum comunica mall-urile pe Facebook”, wall-street.ro, available on-line at <http://www.wall-street.ro/slideshow/Real-Estate/91365/Cum-comunica-mall-urile-pe-Facebook.html>, accessed on 25.08.2010.
43. Onlinecollege.org (2009). 100 Ways You Should Be Using Facebook in Your Classroom, <http://www.onlinecollege.org/2009/10/20/100-ways-you-should-be-using-facebook-in-your-classroom/>, accessed at 10.01.2011.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

ICT and digital sources challenges in the university library: The case of King Khalid University library

S.Q. Al-Khalidi Al-Maliki

King Khalid University, Abha, Saudi Arabia

ABSTRACT: The main objective of this research study is to evaluate the electronic readiness of King Khalid University (KKU) library in order to encourage and assist the KKU library in the use of Information and Communication Technology (ICT) and a digital repository in such a way that will maximise their competitive advantage. Effective use of ICT across all sectors of the economy can act as a driver for increasing competitiveness. Relatively low levels of ICT usage by libraries outside the ICT sector is a contributing factor to the failure of organisations to keep up with productivity growth rates required to succeed. While most countries currently have a strong ICT sector and the potential to do well in newly emerging ICT related industries, performance is far less impressive when it comes to the use of ICT by libraries within non-ICT sectors of the economy. This paper seeks to identify e-readiness and to highlight the weaknesses of implementing ICT within the KKU library. The study concludes with by suggesting an e-readiness model that is appropriate for assessing the existing digital resource and ICT infrastructure.

Keywords: Information and Communication Technology, digital library, e-books, digital data, communication

1 INTRODUCTION

Information and Communications Technology (ICT) is one of the key factors for strengthening growth inconsistencies across countries. ICT covers the acquisition of equipment and computer software. It has three components: information technology equipment (computers and related hardware), communications equipment and software.

The proliferation of technology in the production of ICT goods and services can contribute to faster growth in ICT-producing sectors. The use of ICT helps organisations and people to reduce costs, enhance their productivity and increase their overall efficiency. Moreover, greater use of ICT can contribute to network effects such as lower transaction costs, higher productivity of knowledgeable workers and more rapid innovation, which in turn will improve the overall efficiency of the economy (Al-Maliki, 2013a).

The rapid rate of Internet penetration throughout the world, coupled with dramatic advances in IT use in business and industry, is helping to create an extensive literature base on various aspects of 'e-business' and 'e-commerce', as well as a special interest in 'e-readiness', both here in Saudi Arabia and overseas. The literature has shown that the increased knowledge intensity of economic

activities in almost all industrial countries have contributed to an accelerated interest in e-venues for growth in developing countries. National and international institutions alike appear to be focusing on e-potential for growth in private and public sectors, and almost all developing countries are now mounting national IT development plans (Al-Maliki, 2013a; Choucri, et al., 2003).

E-readiness is a relatively new concept that has been given impetus by the rapid rate of Internet penetration throughout the world, as well as the dramatic advances in IT use in business and industry. The e-readiness concept originated with the intention to provide a unified framework for evaluating the breadth and depth of the digital divide between more and less developed, or developing countries, during the latter part of the 1990s (Mutula & Brakel, 2006).

Measuring access to and the use of ICT is referred to as 'e-readiness', which is the status or quality of readiness of a society or economy for using electronic technology (InfoDev, 2005). High levels of e-readiness allow enterprises to transact business electronically, thus achieving quicker turn-around times, faster delivery of services and enhanced product choices (Mutula & Brakel, 2006).

Information networking for digital integration is becoming of increasing importance for national

and international organisations concerned with development. E-readiness assessments for various countries are associated with the investigation of their state of readiness for such integration. E-readiness assessments help to understand the problems that need to be resolved in order to avoid a digital divide from world development. Such assessments lack standard policies that provide unified measures for easing evaluations, support analysis and comparisons, and that can help in diagnosing problems and deriving solutions (Bakry, 2003).

Since the development of the first e-readiness tool, several others have emerged through the efforts of development agencies, research organisations, academia, business enterprises and individuals. Depending on the objective for assessment, a model is chosen for assessment. The wide range of indicators can be classified according to the following main groups:

- Network access.
- Network learning.
- Network society.
- Network economy.
- Network policy.

This paper presents an electronic readiness assessment framework for King Khalid University (KKU) library's digital access implementation. The framework provides a set of assessment perspectives, including the strategic context for a digital or electronic library, ICT development and policy goals, supply, demand and perceptions. Given a specific assessment scenario, these perspectives are selected and corresponding specialised components address the resulting information needs. To illustrate this process, we show how a concrete readiness instrument can be developed to support Central Library digital access planning for KKU.

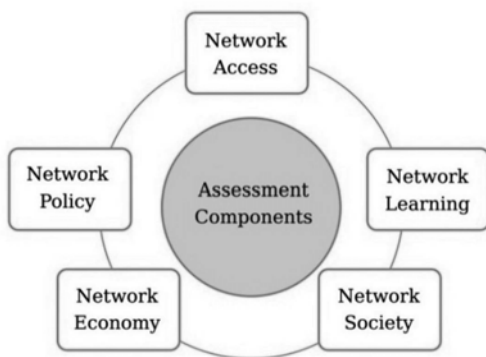


Figure 1. E-readiness assessment components.

2 BACKGROUND STUDY

In South Africa, Mutula et al. (2006) studied the assessment of e-readiness tools with respect to information access. The paper argues that information is a key component in the e-readiness equation. However, it is not given much emphasis in existing e-readiness tools, but is instead subsumed under ICT. The paper proposes a new e-readiness integrated tool that emphasises information access. Furthermore, the researchers observe that it is becoming increasingly difficult to participate adequately in today's global economy without proper e-readiness assessment (Mutula, 2006).

In addition Liljandera et al. (2006) investigated factors affecting customers' attitudes towards these self-study technologies and their adoption behaviour. An empirical study was conducted on the technology readiness of customers of a European airline who had access to an Internet check-in option. The article concluded with a discussion of the validity of the technology readiness construct (Liljander, 2006).

Jaafar et al. (2007) discussed the e-readiness of Malaysian constructors. The authors propose a technology readiness index, a measurement for assessing the readiness of construction technology. The findings are discussed in relation to IT changes that the Malaysian Government is attempting to impose on the country and on the construction industry. The latter, which has always been a follower as opposed to a technology leader, helps to explain the lower level of innovation compared to optimism.

Koh et al. (2008) propose assessing the readiness of a government organisation in order to transform itself into a provider of fully integrated e-government services. They identify major components of e-government and discuss how it can evolve from a simple website into a fully integrated portal that delivers services to the public. Their paper discusses three levels: strategic, system and data.

Nabavi and Davidrajuh (2009) discuss an e-readiness assessment model for evaluating the e-readiness of ICT companies in Iran. The paper discusses an e-readiness model consisting of dimensions and indicators that are selected via a multilateral survey of existing frameworks and models of nations and SMEs in the area of e-readiness assessment. This research shows that two dimensions, networked applications and services, are at a low level among Iranian ICT companies, whereas the electronic infrastructure dimension is at the highest level. This study aimed at designing a model for measuring the e-readiness of ICT companies in Iran.

In terms of the Saudi IT context and the environment of the KKU library, a huge volume of literature has discussed several aspects of ICT. The current researcher wrote on the strengths and

weaknesses of ICT in Saudi Arabia in general, as well as the factors affecting e-governance adoption in this kingdom (Al-Maliki, 2014a, 2013a). In addition, he investigated the current status of KKU libraries, in particular, the use of information resources in the KKU Central Library, and suggests a plan for revitalising academic e-resource sharing (Al-Maliki, 2013b).

The present author also investigated the possibility of implementing cloud storage as an aspect of digital library resource sharing by using a private cloud environment in the KKU deanship of the library (Al-Maliki, 2014b). Moreover, the researcher studied the importance of WiMax equipment in terms of resource sharing within KKU libraries in order to better share e-resources among university staff and students (Al-Maliki, 2015).

3 OBJECTIVES AND ELEMENTS

There are various reasons for the increased stimulus among countries in terms of assessing their e-readiness status. By and large, countries are striving to become inclusive global information societies, where all persons without distinction can be empowered to create, receive, share and utilise information for their economic, social, cultural and political development. Despite the importance of e-readiness assessments, existing e-readiness tools have failed to address the issue of information access adequately. Assessment tools adopt different definitions for the concepts of e-readiness and therefore different ways to measure it, resulting in a variety of assessment, analysis and benchmarking reports with varying degrees of detail.

The need for e-readiness measurement tools is focused on ICT, business, policy and legislative frameworks, and underscore the information access factor. Furthermore, whereas e-readiness research is increasingly populating development, IT and business literature, little development has occurred within the information science discipline. Finally, most e-readiness studies have been confined to macro assessments and have ignored sectorial-level environments.

The Deanship of Library Affairs, established at KKU in 1998 as one of the deanships in the university, serves students, faculty members, lecturers and teaching assistants, supervises the Central Library and its branch libraries, and secures various sources of information. All the colleges libraries inside and outside the city of Abha currently fall under the supervision of the Central Library. The Deanship of Library Affairs at KKU serves as a gateway to all KKU libraries. The Central Library is the most powerful and quickest source of information for academic, scientific and cultural inquiries. The Central Library and the college

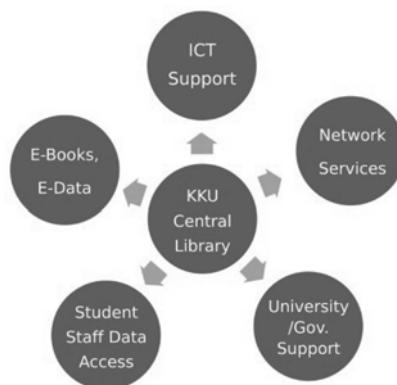


Figure 2. KKU Central Library e-readiness measurement tools.

libraries supply various sources of information. The problem is, however, that many resources and e-books are not being utilised properly due to a lack of ICT, network services, data access problems and security issues. It is essential that these barriers be removed and the strategy of electronic data services be improved.

4 E-READINESS PROBLEM ANALYSIS

The assessment of the e-readiness of the Deanship of Library Affairs was divided into nine segments: ICT facility and structure, Central Library infrastructure and facilities, availability of digital data, network connection availability, digital library users and authenticated persons, availability of funds and utilisation, data security and protection measures, user policy and support, and library facilities and service availability. Furthermore, the assessment was made using several units in each segment. The results of the assessment are recorded in the table below.

The assessment method was divided into three sectors, i.e., strength, weakness and not available. As shown in the Central Library assessment table, availability and strength was strong, but other segments were weak in terms of e-readiness as it pertained to the deanship of the KKU library.

The KKU library's e-readiness was conducted with the aim of leading to ICT, as well as to being able to effectively recognise, gather, organise, process and distribute information for making effective administration decisions. Additionally, the importance of the e-readiness assessment will create effective and economic digital sharing within the networked infrastructure among other community colleges that are part of KKU.

The proposed assessment of e-readiness was developed after reviewing the related literature and

Table 1. E-readiness assessment table.

S. No	Assessment aspects	Measures	Strength	Weakness	Not available
1.	ICT facilities and structure	Web portal Staff services Student services Central Library Digital library services E-learning services		*	* * * * *
2.	Central Library infrastructure and facilities	Data availability Digital data Student access Staff access Faculty access Teaching assistant E-learning support	* * * * * *		*
3.	Availability of digital data	Books E-books E-magazine E-data E-study materials		* * * *	*
4.	Network connection availability	Wireless Wi-Fi Mobile access LAN connection WiMax Broadband support		* *	* * * *
5.	Digital library users and authenticated persons	Students Staff Faculties Teaching assistants Administrators Public	* * * *	*	*

Number of Employees in KKU Library

Men: 30, four of whom are specialised in information systems.

Women: 35, one of whom is specialised in library and information sciences.

6.	Availability of funds and utilisation	Subscription to national and international digital database Data access Digital books	*	*	*
7.	Data security and protection measures	Student access security Staff access security Administrator access security Faculty access		* * * *	
8.	User policy and support	Students Staff Faculties Administrators Public		* * *	* *

(Continued)

Table 1. (Continued)

S.No	Assessment aspects	Measures	Strength	Weakness	Not available
9.	Library facilities and service availability	Reading	*		
		Borrowing books	*		
		Searching e-databases	*		
		Copying		*	
		Printing		*	
		Translation			*
		Internet services		*	
		Security gates		*	

Number of PCs in KKU Library

Number of PCs in the library: 30

However, 22 libraries in KKU are without any IT or PCs

investigating the primary parameters and criteria that could be used for evaluating the e-readiness of the KKU library. The proposed assessment of e-readiness employed will have special significance for the KKU library in Saudi Arabia; within the context of being a developing country, the assessment will assist in making clear implementing the ICT in our libraries at KKU.

5 E-READINESS ASSESSMENT MODEL FOR KKU LIBRARY

The proposed e-readiness tool is designed around six major segments: ICT readiness, user readiness, digital data readiness, network structure readiness, fund support readiness and data security readiness. Around each of these segments, a set of rules can be created to measure the degree of e-readiness within the library deanship, with information access taking its rightful place as a key determinant for effective digital sharing.

5.1 ICT readiness

As shown in the above table, the analysis report notes that there is not enough ICT readiness within the KKU library. All points indicate only its weaknesses, as well as the non-availability of facilities in the library deanship. Hence, there must be a greater focus on constructing effective e-readiness. If the objective is improving the e-infrastructure of the library, then the focus should be on digital data, network connectivity, hardware and software. Here, e-readiness equals computers and access; computer hardware and network access are required to be e-ready and to bridge the digital divide and university administration and ICT

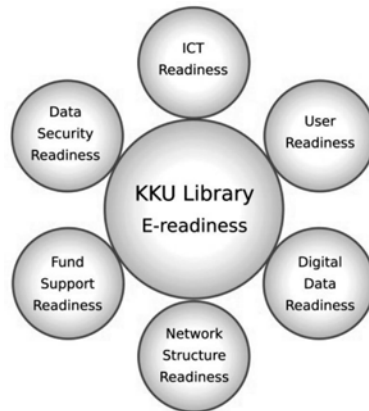


Figure 3. Proposed structure of KKU e-readiness assessment tool.

administrator initiatives should be employed as solutions.

5.2 User readiness

The aim of creating an effective digital library is to provide students and the staff community with digital books, academic and scientific references and e-journals. Here, e-readiness requires basic literacy, a usage policy and accessibility, as well as other social issues to be addressed first; computers are useful, but nothing will render a society e-ready and bridge the digital divide until basic literacy, deficiency and usage policy, and the unrestricted accessibility of digital mediums have been addressed. As shown in the above table, there is a lack in point of user supporting policy. Therefore,

we need to establish a stronger user support policy for unrestricted and constant communication between the library deanship and the user.

5.3 *Digital Data Readiness*

The digital data focus should be on readable process reengineering and a faster and more transparent means of delivering library services to students and staff. Here, e-readiness is represented by the number of computers, access to them and the effective use of said computers. Hardware and access alone are not sufficient for true e-readiness; extensive training programmes must be provided for library staff, locally relevant content must be made available and a local ICT sector established.

5.4 *Network structure readiness*

Effective access to e-networks requires an infrastructure system consistent with present conditions, coupled with reliability as it pertains to all supporting services. Both sets of access conditions must be stable, with faster access to more information, improved communication and collaboration, as well as more convenient access to software tools. We should concentrate on factors affecting e-readiness performance that are largely socially constructed, including norms, values, education, training and modes of financial activity. The e-readiness network structure, with both synchronous and asynchronous communications, can include natural voice and video, so that all participants can see and hear one another. In creating effective network structure support for e-readiness, we should implement advanced data connectivity like 4G, Wi-Fi, WiMax and Long Term Evaluation technology (LTE).

5.5 *Fund support readiness*

Funds constitute an important factor for the success of the implementation of the strategy. An estimation of funds is needed for the implementation of major e-readiness strategy initiatives. The e-readiness strategy can specify a number of fund sources. There is no proper or consistent funding source for e-readiness projects in terms of improving the KKU library. As e-relative programmes cover an extended period of time, it is essential that such programmes are funded continuously in order to keep them on track. Many projects that have failed have done so due to a lack of funding on behalf of university administrations and governments.

5.6 *Data security readiness*

Categorisation in security should also consider vulnerabilities and provide information about known

threats to information systems and data transfer applications. Security at all levels is essential for achieving a secure environment; the probability of simultaneous security breaches at all levels is less likely in digital information sharing. Regarding the legal foundations for digital development in library data, the challenges are also getting tougher. As technologies advance and become more intrusive, usage and data volumes increase and become more sophisticated.

6 BENEFITS OF THE PROPOSED SYSTEM

In academic institutions, libraries are becoming ever larger. Implementing assessments for e-readiness is necessary to highlight any weaknesses in terms of e-readiness. Therefore, implementing e-readiness assessment will enable us to gain the following benefits:

- The assessment system measures the current level of ICT structure.
- The assessment system identifies the current level of understanding and provides the opinions of students, staff and society for effective e-readiness.
- The assessment system provides information on the available ICT infrastructure, network structure, resources, applications and services.
- The proposed e-readiness assessment framework has the potential to assist the university's administration, the ICT unit and other departmental libraries in KKU, especially in the library deanship, to make informed ICT investment assessments.
- The assessment provides information on the existing regulatory and legal environment.
- It provides implementation suggestions for ICT infrastructure.
- Extending digital resources to other universities/colleges.
- Improving network structure.
- Extending the library service to students, staff and government.
- The assessment system enables information about primary objectives with respect to national level commitments such as e-governance related to community/societal development.
- The proposed system enables creativity and sharing, as well as skills development resource boundaries within the student community.
- The proposed e-readiness will improve ICT quality and reliability.
- The system will enable strong digital resource sharing, and library data availability.
- E-readiness assessments will enable the use of KKU library digital resources, it is vision, strategy, priorities and expansion.

The elements highlighted above demonstrate that factors can contribute either positively or negatively to KKKU library e-readiness. Managers' perspectives and enthusiasm can generate ideas and lead to positive practices. It is argued that the past experience of managers and how relevant their scientific interests and backgrounds are to ICT and libraries should be considered as important constraints for assigning library directorships. The predicament here is that large numbers of resources and e-books are not being properly utilised, due to the lack of a clear management perspective. In addition, the absence of a solid business strategy and the lack of ICT, network services, data access difficulty and security also contributes to the weaknesses of implementing ICT within the KKKU libraries. It is essential that these barriers be removed and strategies for electronic data services be improved. Therefore, it is of a great importance at this stage to determine these barriers and to thereafter develop an appropriate electronic data services strategy.

7 CONCLUSION

This research study presented a structure that has the potential for assisting the KKKU library, especially in digital data availability, as well as to make informed ICT investment decisions that will enable electronic data to increase the electronic resource sharing structure. Our review showed that none of the major readiness assessment frameworks for digital or e-libraries in Saudi Arabia covers these different aspects or provides the necessary information for effective strategic planning for digital resource sharing. This study thus provides a benchmark for a library's e-readiness status, which will help to discover weaknesses and areas that need more attention. The sample and data obtained from the Deanship of Library, as well as from library staff and students, helped us to obtain insight into digital resource sharing information and issues related to digital library requirements. Our proposed e-readiness assessment structure suggests that a six-level e-library readiness model is more appropriate than the existing digital resource sharing system. Our proposed structure creates interaction among students, staff, administration and society, and also assists them to benefit from government services in a transparent and efficient manner. Taking the initiative on e-readiness, the proposed structure will involve preparing the student community to familiarise itself with the importance of digital information sharing, and doing so in an effective manner. Additionally, our e-readiness assessment system is oriented towards promoting the use of ICT, enhancing ICT infrastructure, sharing digital

mediums and creating the necessary regulatory framework for retaining next generation networks. Perhaps this study will engender more discussion on framework strategy components, for example, addressing the issue of why each component is considered an obstacle. Furthermore, it may lead to in-depth discussions on the best way for removing these barriers.

REFERENCES

- Al-Maliki, S.Q. Al-Khalidi, (2013a), Information and Communication Technology (ICT) investment in the Kingdom of Saudi Arabia: Assessing strengths and weaknesses, *Journal of Organizational Knowledge Management*, 1–15.
- Al-Maliki, S.Q. Al-Khalidi, (2013b), A new plan for King Khalid University (KKU) Central Library to revitalise academic e-resource sharing, *International Research: Journal of Library & Information Science (IRJLIS)* 3, no. 4, 596–602.
- Al-Maliki, S.Q. Al-Khalidi, (2014a), Analysis and implementation of factors affecting e-Governance adoption in the Kingdom of Saudi Arabia, *International Journal of Strategic Information Technology and Applications (IJSITA)* 5, no. 1, 20–29.
- Al-Maliki, S.Q. Al-Khalidi, (2014b), Implementation structure of Cloud storage in digital library resource sharing using private cloud environment: A new approach in the Deanship of the Library, King Khalid University, Saudi Arabia. In *Proceeding of the 2014 International Conference on e-Commerce, e-Administration, e-Society, e-Education, and e-Technology—Fall Session (e-CASE & e-Tech 2014 – Fall Session)* to be held in Tokyo, Japan, Nov. 12–14, 2014.
- Al-Maliki, S.Q. Al-Khalidi, (2015), The significance of WiMax on resource sharing among Saudi Universities: A study of King Khalid University, Kingdom of Saudi Arabia, *International Journal of Information and Education Technology (IJIET)* 5, no. 12, 961–964.
- Bakry, S.H., (2003), Toward the development of a standard e-readiness assessment policy, *International Journal of Network Management* 13, issue 2, 129–137.
- Choucri, N., Maugis, V., Madnick, S., Siegel, M., (2003), Global e-Readiness—for what?, *Report of the Group for Globalization of e-Business*, Center for eBusiness at MIT, Sloan School of Management.
- InfoDev, bridges.org, (2005), E-Ready for What? E-Readiness in developing countries: Current status and prospects toward the Millennium Development Goals, Cape Town, [Online] Available at: http://www.infodev.org/infodev-files/resource/Infodev_Documents_3.pdf
- Jaafar, M., Rashid, A., Ramayah, A. T., Saad, B., (2007), Integrating information technology in the construction industry: Technology readiness assessment of Malaysian contractors, *International Journal of Project Management*, 25, 115–120.
- Koh, C.E., Prybutok, V.R., Zhang, X., (2008), Measuring e-government readiness, *Information & Management* 45, 540–546.

- Liljander, V., Gillberg, F., Gummerus, J. Riel A.V., (2006), Technology readiness and the evaluation and adoption of self-service technologies, *Journal of Retailing and Consumer Services*, 13, 177–191.
- Mutula, S.M., Brakel, P.V. (2006), An evaluation of e-readiness assessment tools with respect to information access: Towards an integrated information rich tool, *International Journal of Information Management* 26, 212–223.
- Nabavi, A., Davidrajuh, R., (2009), Designing an assessment tool for measuring e-readiness of Iraninan ICT companies, *Issues in Information Systems*, 10, no. 2, 175–184.

An exploratory evaluation of the awareness of e-government services among citizens in Saudi Arabia

S.Q. Al-Khalidi Al-Maliki

King Khalid University, Abha, Saudi Arabia

ABSTRACT: Many studies have been carried out on the awareness of using e-governance services, but very few studies have looked at this topic in the context of Saudi Arabia. This study aims to identify the ways in which the awareness of using e-governance services can be increased among citizens in the southern region of Saudi Arabia, namely the cities of Abha and Jazan, and to explore factors that hinder public organisations from benefiting from the provision of e-governance services to the citizens. This studying can be generalised among Saudis in other cities of Saudi Arabia. The sample data were collected from service offices, citizens, university students, and government employees. In addition, in order to gain an insight into the ways of using e-services, 45 randomly selected citizens were interviewed in the vicinity of government agencies in the two cities, in order to ascertain their point of view regarding the issues and factors that affect awareness of the use of e-government services. This paper analysed the findings of those interviews to provide various recommendations with regard to the usage policy of e-governance, facility and infrastructure, and the extent of e-government awareness among Saudis. It is hoped that this study will increase the awareness of the benefits of using e-governance in the day-to-day life of Saudis and public and private organisations.

Keywords: e-services awareness, e-Government, information and communication technology, IT investment, Saudi Arabia

1 INTRODUCTION

With rapid developments in Information Technology (IT), the effects of Information Systems (IS) on employee performance in organisations are evolving. Organisations are introducing computer technology and developing their own IS for more efficient management. The increasing utilisation of IS may encourage employees to use it to help them perform tasks and manage work. This has resulted in the rapid development of the electronic government (e-government) concept. The concept refers to the use of IT/IS by the government to provide citizens and organisations with more convenient access to government services. E-government provides government-related information and services to citizens via Internet and non-Internet applications.

The main purpose of e-government is to build a government that exists everywhere and is ready to serve at any time. Through the use of different information equipment, e-government allows enterprises and the public to receive related services at any time and place. However, the implementation of such new and innovative policies requires consensus among most of the personnel within the concerned organisation. Therefore, employee

performance can be regarded as an outcome of e-government. Accordingly, employee performance via IS use becomes an important theme in the context of e-government (Luarn and Huang, 2009).

By using Information Communication Technology (ICT) tools and applications, the Internet, and mobile devices to support good governance, government agency and department initiatives strengthen existing relationships and build new partnerships within civil society. These are known as e-government initiatives. As with e-commerce, e-government represents the introduction of a great wave of technological innovation, as well as government reinvention. It represents a tremendous impetus to move forwards in the 21st century with higher quality, cost-effective government services and a better relationship between citizens and governments (Ndou, 2004, Fang, 2002). Many government agencies in developed countries have taken steps towards Web and ICT use, thereby making all local activities coherent on the Internet, widening local access and skills, opening up interactive services for local debates, and increasing the participation of citizens in the promotion and management of territory (Ndou, 2004, Graham and Aurigi, 1997).

In this research paper, the background literature will be presented following the introduction, then, in section 3, the research problem statement will be discussed. A research objective and methodology will be presented in section 4, followed by the sampling and data collection procedure and interview questions in section 5. Next, in section 6, the findings and discussion will be presented, followed by the implications and conclusion in sections 7 and 8.

2 BACKGROUND LITERATURE

This paper will present a review of the available literature on the topic of e-government adoption. Previous studies by this author (Al-Maliki, 2014) and Abanumy, et al., (2005) explored three main topics: website accessibility guidelines, website accessibility tools, and the implication of human factors in the process of implementing successful e-government websites. These studies examined the issues that make a website accessible and explored the importance placed on web usability and accessibility with respect to e-government websites. They also briefly examined accessibility guidelines, evaluation methods, and analysis tools. Alshawi and Alalwany (2009) discussed the ways in which e-government, in both theory and practice, has proven to be important and complex. The paper is part of research efforts to develop a rigorous and holistic evaluation framework for e-government systems. The main aim of this article was to develop evaluation criteria for an effective, adaptable, and reflective assessment of e-government systems from the perspective of citizens. The criteria can also be used as a means for providing valuable feedback for planning future e-government initiatives. Chatfield and Alhujran (2009) provided an insight into the current state of the development of Arab e-government programmes. A cross-country comparative analysis of e-government websites and portals involving 16 Arab countries assessed their relative developmental stages in terms of capability of delivering e-government services. The results revealed a wide digital divide among the Arab countries studied, particularly in terms of the development and capacity to deliver advanced e-government services. These results have important implications for developing countries in terms of managing both economic and non-economic resources effectively for successful e-government development. Luarn and Huang (2009) investigated the implications and consequences of government employee performance via IS. A multiple regression method was used to investigate factors that influence employee performance. The results

indicated that three factors affect performance: task-technology fit, computer self-efficacy, and utilisation. Utilisation was found to have the most positive effect on performance. In addition to verifying prior empirical findings, this study outlined the factors that influence employee performance and IS development in the context of e-government.

In addition, Al-Solbi and Al-Harbi (2008) explored key e-government policies and factors that contributed to the success of e-readiness assessment from the perspective of some of the public and private organisations in Saudi Arabia. With this aim, a questionnaire was developed and distributed, and semi-structured interviews were conducted with ICT managers in specific organisations. The findings are very important since they indicate that both Saudi and organisational leadership are important aspects of ICT infrastructure in Saudi society. Al-Jaghoub, et al., (2010) conducted a study on awareness and acceptability of using e-government services entitled, "Evaluation of Awareness and Acceptability of Using e-Government Services in Developing Countries: the case of Jordan". Their study showed that awareness of e-government in Jordan was not at the required level. They identified factors influencing the usage of e-government services, such as cultural resistance and lack of trust in online services.

In "Challenges of e-government services adoption in Saudi Arabia from an e-ready citizen perspective", Alshehri and Drew (2010) identified technical, ICT infrastructure, organisational, security, social, and culture barriers and made recommendations to overcome them.

In another study, Alshehri, et al., (2012) investigated some obstacles and challenges in the adoption of e-government services in Saudi Arabia. They identified many important factors that directly affect the adoption process. Also, they made recommendations to help the public sector and government organisations to improve their electronic services.

A study (Al-Maliki, 2013) identified the impact of ICT investment in Saudi Arabia and the role of government through a series of "Five-year plans". It also outlined and analysed current IT use and development in Saudi Arabia, as well as the factors affecting economic growth. This research also assessed strategies and policies related to ICT and investment in ICT in Saudi Arabia, and discussed the role of public and private organisations, as well as educational institutions at all levels (Al-Maliki, 2013). Despite progress in the use of ICT in Saudi public and private organisations, there are still challenges to overcome

before ICT becomes a viable part of Saudi life. Moreover, another study (Al-Maliki, 2014) determined the factors affecting e-governance adoption in Saudi Arabia. It was found that the main barriers were a lack of skills and human resources, low computer literacy and training capacity, and English language difficulties. The researcher proposed a conceptual architecture for e-portal services. The proposed recommendations addressed the need to understand the adoption of e-government and to help citizens use the available services. The proposed model reiterated that citizens should be helped until they accurately understand the functioning of e-government applications.

Basamh, et al., (2014) studied the adoption and implementation of e-government in Saudi Arabia. They explored current practices, obstacles, and challenges affecting the improvement of e-services from the perspective of society. They found that infrastructure costs, computer literacy, privacy issues, accessibility and availability, and trust issues were major obstacles to the implementation and adoption of e-government in Saudi Arabia.

3 RESEARCH PROBLEM STATEMENT

E-government is a prominent concept today in both popular and academic discussions on governance reform. Studies on awareness of using e-services in Saudi Arabia show that a number of challenges have hindered the reach and impact of e-government. Several social, economic, and literacy barriers constrict the scope of transformation and restrict the ability of policy-makers to effectively use new e-government technologies. There are many potential barriers to the implementation of e-governance:

- ICT skills.
- Technology factors.
- Cultural differences.
- Integration technology.
- E-governance application software.
- Government support.
- Trust and security.
- Digital divide.

However, the obstacles listed above differ from one country to another according to usage, facilities, and culture. These obstacles and constraints require urgent solutions. In the Saudi context, the main obstacles are related to socio-cultural, organisational, and technical factors (Al-Maliki, 2013) and lack of awareness and trust. A review of the literature revealed that there is limited empirical research on e-government services from

the citizens' perspective. According to Al-Maliki (2013) cultural differences can result in limitations in IS implementation. For example, different views on logic and reasoning, and limitations in language use can create a barrier to effective communication and understanding.

The Saudi government established a gate to e-government services, called "Saudi", which is an e-government portal. It is a national portal for accessing e-government services in Saudi Arabia from anywhere, available to citizens, expatriates, companies, and visitors. E-services are delivered to users in a highly efficient manner. The e-government programme ("Yesser") launched its national portal on 27 March 2006. E-services are accessible through the "Saudi" portal, either by integration with other government agencies or via links to those agencies or their services on the portal. The Saudi portal has around 2035 e-services of many public and private organisations in Saudi Arabia (Saudi official portal, 2016). However, Saudis need to be made aware that e-government offers a number of potential benefits. Owing to the lack of awareness, some agencies offer e-services, for example, bookshops, and some have established offices to help the citizens.

4 RESEARCH OBJECTIVES AND METHODOLOGY

The objective of this paper is to evaluate the extent of awareness among Saudis in using e-government services. It also aims to identify the main factors that may inhibit the use of e-services in government organisations.

The research adopted descriptive and analytical approach, involving the study of the use of e-government services and interviews with citizens. The interviews allowed more in-depth and, in some cases, broader responses from the respondents than would have been the case with other fact-finding approaches. Interviews were carried out with major players in e-government in the cities of Abha and Jazan, located in southern Saudi Arabia, namely the municipalities, e-government agencies, and some departments in the Ministry of the Interior.

A literature review was also conducted on this topic. Information was collected from articles published by other researchers and from current trends in this area. In addition, data was collected through a website content study. Moreover, observations were carried out at some agencies and service offices. This research approach was considered appropriate for analysing citizens' awareness of using e-government services in Saudi Arabia.

The interview questions were open questions related to various aspects of using e-government services. The following issues were covered: extent of awareness about the use of e-government services in Saudi, the main factors hindering awareness of the use of e-governance services over the Internet, the principal reasons behind Saudis not being very familiar with e-government services, the main factors that could lead to resistance or failure of using e-government services, whether or not the e-government service providers were satisfied with the extent of citizens' awareness, solutions to the lack of awareness among Saudis, and the appropriateness of e-services websites.

5 SAMPLING AND DATA COLLECTION PROCEDURE

E-government has the power to serve and implement good governance, economic growth, and human development through increased efficiency, accessibility, transparency, and accountability of government operations, leading to improved national performance in all aspects.

For this study, 45 citizens were interviewed and 15 service offices visited to observe how citizens seek help and how e-services are executed. The sample was drawn from southern Saudi Arabia, in particular, the cities of Abha and Jazan, and can be generalised among Saudis in other cities of Saudi Arabia. The study population was defined as citizens, government employees, and university students. The study sample included citizens who worked at the offices or offered services to use the e-services and citizens who needed the services. Citizens who visited e-government agencies to process e-services were selected randomly. The researcher met the respondents over three weeks at different times, at the following locations: passport departments, civil agencies, labour offices, Chamber of Commerce branches, Ministry of Commerce branches, Ministry of the Interior sections, and municipalities. The sample size can be considered adequate for this study.

The Saudi government provides many e-services via its government website portal "Saudi"; an illustration is shown in the table below:

5.1 Interview questions

To achieve the study objectives, literature on the awareness of using e-governance services among Saudi citizens was reviewed. The following questions on this topic were then addressed:

1. What is the extent of the awareness of using e-government services among Saudi citizens?

2. What are the main factors hindering the awareness of using e-governance services over the Internet among Saudi Arabian citizens?
3. What are the principal reasons behind Saudi citizens being less familiar with the use of e-government services within public organisations?
4. What are the main factors that could lead to resistance or failure of the use of e-government services provided by Saudi public organisations?
5. Are the public organisations implementing e-government services satisfied with the extent of citizens' awareness about the e-services they provide?
6. What are the suitable solutions to address the lack of awareness among Saudis?

According to Alshehri and Drew [11], most e-government websites are inefficient and provide just basic and general information about organisations, plus often the data is not updated. Therefore, the following specific questions about e-government service websites were posed to interviewees:

1. What is your overall impression of e-service websites?
2. Do e-service websites provide the information you need?
3. Does the content meet your needs?
4. Are you satisfied with the accuracy of e-service websites?
5. Is the help guide of the website clear?
6. Do the related e-service agencies provide up-to-date information?
7. Is the e-service website user friendly?
8. Are the e-service websites easy to use?

In addition, the interviewees were asked to what extent they agreed/disagreed with the following statements related to the resistance or failure of using e-government services:

6 FINDINGS AND DISCUSSION

E-government supports broad public sector reforms and good governance through the introduction of innovative and sustainable applications of ICT within government administrations, for enhanced interaction with citizens and the private sector. The public sector is increasingly seen as the main element that can bridge the digital divide at national level. Public agencies need to be model users of ICT so that others can follow. The public sector tends to be the biggest provider of local content and it can nurture and foster further development of the local ICT industry.

Table 1. E-government services. (Source: saudi.goc.sa, 2016).

 <p>ICT</p> <p>(71) e-Services</p>	 <p>Economy & business</p> <p>(158) e-Services</p>
 <p>Islamic affairs</p> <p>(110) e-Services</p>	 <p>Labor & employment</p> <p>(109) e-Services</p>
 <p>Housing & municipal services</p> <p>(372) e-Services</p>	 <p>Health & environment</p> <p>(171) e-Services</p>
 <p>Traffic & safety</p> <p>(372) e-Services</p>	 <p>Personal documentation</p> <p>(84) e-Services</p>
 <p>Training, education and culture</p> <p>(579) e-Services</p>	 <p>Travel & tourism</p> <p>(55) e-Services</p>
 <p>Social life</p> <p>(66) e-Services</p>	 <p>Insurance & pension</p> <p>(44) e-Services</p>
 <p>Utilities</p> <p>(132) e-Services</p>	 <p>Transportation</p> <p>(55) e-Services</p>

Many interviewees stated that the main obstacles to adopting e-government services were a lack of trust in and awareness of using online services. In addition, many interviewees indicated that, in their opinion, e-service departments in all organisations should provide better help and support to improve citizens' awareness. The majority of the interviewees suggested that one of the main barriers to the use of e-services in Saudi Arabia was the lack of awareness about how to execute the e-services. Appropriate help and support files should be included within e-service websites.

In some cases, citizens do not understand how to use e-services owing to a lack of awareness. Furthermore, some citizens complained about the difficulty of using e-services because they were not familiar with e-government services.

It seems that a lack of awareness of using e-services within Saudi public organisations is one of the major barriers to e-governance. Many organisations in Saudi look at e-government services as an important way to improve internal operations and provide quick and more efficient operations and better quality services, but they should educate their clients regarding how to use their e-services.

Table 2. Problems arising when accessing e-services.

No	Statement (Factor)	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1	Perceived difficulty or complexity of using e-services					
2	Lack of understanding					
3	Lack of sufficient help					
4	Lack of awareness about using e-services					
5	Lack of organisation help desk support					
6	Inappropriateness of the e-service applications					
7	High cost of using e-services					
8	Lack of website support for using e-services					
9	Using e-services is not considered an important process					
10	Using e-services is not required in our daily life					
11	Other factors					

Saudi public organisations should provide awareness programmes and tools to help users learn about how e-government applications work, as well as train their employees so that they have a sufficient understanding of how the applications work and, in turn, help citizens.

Concerning training, trainers are an important factor in the implementation and integration of computer technology in education. Without sufficient knowledge, it will be difficult to appropriately use the e-services. Appropriate training is a must for all professional e-services users in order to possess the skills and awareness needed to use these applications.

The major obstacle to successful implementation and usage of e-government applications according to this study is the lack of user awareness and motivation. Weak communication between e-government employees and citizens as clients can explain misunderstandings over the use of e-governance applications. Also, there is a lack of awareness of current services and their usefulness.

Owing to inadequate training support from the vendor, public organisations have had to spend hours learning and training their personnel to use different software applications. Owing to the complexity of these applications, many citizens have to develop their own knowledge by using the Google search engine or by contacting a friend to get help to use these services. Saudi public organisations should address this problem by helping people learn to use e-services.

There is also a lack of e-service expertise within organisations, namely suitable help and support files or clips showing how to use the services. It seems that some organisations do not think

training is important enough to affect their services. Most IT failures are due to inadequate training programmes provided either by the system's users or vendors.

The main points of help for citizens to execute e-services are the service office, estate office, friends and colleagues, bookstores, agents working for government and service agencies, and office imaging services.

Other major factors influencing the awareness of using e-services are electronic illiteracy, distrust, fear of making mistakes, absence of an e-mail ID that is required for a user to be able to open an account on the e-service website, lack of awareness of the existence of these services, lack of smart devices, and dependency on colleagues and friends to carry out the e-services.

Lack of updated information on the websites could be a result of the current lack of awareness within e-services management departments. The findings of this study highlighted a number of issues related to e-government applications, appropriate use of e-government applications, availability of up-to-date information, training of users, and provision of support to citizens.

Concerning the e-service websites, the findings indicate that there is a lack of attention given to content quality. The websites do not provide the "Frequently Asked Questions" service, where visitors can get the required information as soon as possible without having to spend a great deal of time searching every part of the website.

Some websites lack contact details ("contact us"), i.e. phone, fax, or email. Some e-service websites are not compatible with specific Internet browsers. In addition, e-service links or icon images

are not always in obvious places on websites' main pages, difficulty of website interface, and no possibility in the e-service website to follow-up the transactions or process executed.

7 IMPLICATIONS

On examining the main findings of this study, it can be seen that e-government service departments should raise the quality of e-service awareness to help the users, i.e. citizens, residents, and visitors, to provide effective and efficient government services and to meet the needs of the beneficiaries of government agencies.

E-government websites should include enough information about their e-services. Also, e-government departments should examine their user satisfaction forms to evaluate the performance of their e-service applications. Public organisations that provide online services should encourage citizens to use their services by educating them about these services. Also, e-service websites should meet the requirements and desires of users and be easy to understand and use. The findings of this study will help government and private employees and citizens gain a deeper understanding of e-governance use. It will create public awareness about the potential of ICT. Citizen access to government information/services must increase to combat the digital divide. Urgent training in ICT-based systems and services can enhance the knowledge and skills of the concerned parties. Increased awareness of the use of e-services will increase IT literacy and help reduce the internal digital divide. This will also help boost the quality of education, health services, and social security. Further, this will play a major role in increasing the capacity for rational distribution of public funds and strengthening a development-oriented and people-centred service-delivery culture. In addition, e-government agencies should give priority to ensuring that Internet channels are safe and reliable in order to build a reliable ICT infrastructure and to avoid breakdown of services.

In summary, e-governance involves the entirety of society and technology—e-governance is now integrating all the services in a single system. Saudi public organisations should analyse their systems on a regular basis to reassess their readiness for technological progress and ongoing changes in the governance system.

According to Table 1, many e-service agencies in Saudi Arabia post relevant information online in an organised and easy-to-access manner for other government agencies, businesses, and citizens. Also, relevant transactions between government agencies and private sector businesses and citizens can now take place online.

8 CONCLUSION

In the last few years, there has been considerable infrastructure development in IT, and we need to think about where we will be and where we want to be with regard to e-governance infrastructure to ensure connectivity between the public and government.

The researcher adopted a descriptive survey method involving face-to-face interviews, considered appropriate for collecting sufficient qualitative data on the awareness of using e-services in southern Saudi Arabia. In this current research, many conclusions were drawn, offering the possibility of reshaping the public sector's activities and processes, building relationships between citizens and government, enhancing transparency, increasing government capacity, and providing infrastructure facilities.

The qualitative analysis provided a general overview of current awareness of using e-services by examining the responses of 45 Saudi citizens. The interviews and observations gathered a broad range of ideas regarding how citizens execute e-services. On the whole, the interviewees gave very similar answers about their difficulties and the need to understand the process and overcome specific obstacles in using e-government services. The findings indicate that there is a poor understanding of the use of e-services. There has been tremendous infrastructure development in IT and e-governance in Saudi public organisations to enhance connectivity between the public and government organisations. This is an important finding that reflects the fact that ICT is very heavily invested in Saudi public organisations, but more effort is required with regards to the support and help needed for using e-government services via the Internet.

This study also showed that there is a poor understanding of the process of using e-services and that Saudi citizens' awareness is limited in some areas. There was an evident lack of management support, which many felt hindered the help and support given to citizens, and there was a definite lack of understanding of e-government services, namely the functioning of public organisation websites. In addition, many organisations, unfortunately, showed a lack of willingness to help users. E-government agencies working for many public and private organisations must overcome a number of obstacles to achieve the growth necessary for e-government services to be delivered to all Saudi communities.

Finally, this study placed demands on public organisations that need more attention, and it may lead to in-depth further discussion on the best way to increase awareness of e-government services among Saudis and studies on some aspects of e-government services.

REFERENCES

- Abanumy, A., Al-Badi, A. and Mayhew, P., (2005), E-Government Website Accessibility: In-Depth Evaluation of Saudi Arabia and Oman. *The Electronic Journal of e-Government*, 3(3), pp. 99–106.
- Al-Jaghoub, S., Al-Yaseen, H. and Al-Hourani, M., (2010), Evaluation of Awareness and Acceptability of Using e-Government Services in Developing Countries: The Case of Jordan. *Electronic Journal Information Systems Evaluation*, 13(1), pp. 1–8.
- Al-Maliki, S.Q. Al-Khalidi, (2013), Information and Communication Technology (ICT) Investment in the Kingdom of Saudi Arabia: Assessing Strengths and Weaknesses. *Journal of Organizational Knowledge*, 2013, pp. 1–15.
- Al-Maliki, S.Q. Al-Khalidi, (2014), Analysis and Implementation of Factors Affecting e-Governance Adoption in the Kingdom of Saudi Arabia. *International Journal of Strategic Information Technology and Applications*, 5(1), pp. 20–29.
- Alshawi, S. and Alalwany, H., (2009), E-government evaluation: Citizen's perspective in developing countries. *Information Technology for Development*, 15(3), pp. 193–208.
- Alshehri, M., Drew, S. and Alfarraj, O., (2012), A Comprehensive Analysis of E-government services adoption in Saudi Arabia: Obstacles and Challenges. *International Journal of Advanced Computer Science and Applications*, 3(2), pp. 1–6.
- Alshehri, M. and Drew, S., (2010), Challenges of e-Government Services Adoption in Saudi Arabia from an e-Ready Citizen Perspective. *World Academy of Science, Engineering and Technology*, 66, pp. 1053–1059.
- Al-Solbi, A.N. and Al-Harbi, S.H., (2008), An exploratory study of factors determining e-government success in Saudi Arabia. *Communications of the IBIMA*, 4, pp. 188–192.
- Basamh, S.S., Qudaih, H.A. and Suhaimi, M.A., (2014), E-Government Implementation in the Kingdom of Saudi Arabia: An Exploratory Study on Current Practices, Obstacles & Challenges. *International Journal of Humanities and Social Science*, 4(2), pp. 296–300.
- Chatfield, A. and Alhujran, O., (2009), A cross-country comparative analysis of e-government service delivery among Arab countries. *Information Technology for Development*, 15(3), pp. 151–170.
- Fang, Z., (2002), E-Government in Digital Era: Concept, Practice and Development. *International Journal of the Computer*, 10(2), pp. 1–22.
- Graham, S. and Aurigi, A., (1997), Virtual Cities, Social Polarisation, and the Crisis in Urban Public Space. *Journal of Urban Technology*, 4(1), pp. 19–52.
- Luarn, P. and Huang, K., (2009), Factors Influencing Government Employee Performance via Information Systems Use: an Empirical Study. *Electronic Journal of e-Government*, 7(3), pp. 227–240.
- Ndou, V., (2004), E-Government for developing countries: Opportunities and Challenges. *The Electronic Journal of Information Systems in Developing Countries*, 18(1), pp. 1–24.
- Saudi official portal. Title. SAUDI [online] Available at: <<http://www.saudi.gov.sa/>> [Accessed: 20 January 2016].

The role of social media in creating new knowledge for graduate students at King Abdulaziz University in Jeddah, Saudi Arabia

Suzan Ahmad Al-Afghani & Hessah Mouner Albogami

Department of Information Science, King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT: This study aims to investigate the role of Social Media in creating new knowledge for graduate students at King Abdulaziz University in Jeddah (Saudi Arabia) as well as find out the degree of using these media and how they use it in accessing new knowledge. In order to detect these things; a questionnaire was designed and distributed on female students of King Abdulaziz University. We use the correlative descriptive survey method to solve the problem of this article. The study found that the degree of graduate students' employment of social media for access to acquire knowledge was medium degree. Perhaps a lack of electronic communication through social media between some professors and students contributed to decline the employment of social media for access to knowledge.

Keywords: Social Media, Knowledge, E-Learning, Facebook, Twitter

1 INTRODUCTION

Recently, the academic interest is increased and developed with issues of networks and virtual communities, which contributed to the creation and development new concepts in the field of learning and teaching such as e-learning, distance learning, virtual universities and online learning which promoted the concept of lifelong learning. This principle depends on the idea that learning is not limited to the period of childhood or in classroom, but exceeds all ages at any place. Among the most prominent and controversial of these issues are social networks.

As social networks helped in solving a lot of educational problems including the lack of humanitarian aspect in e-learning through human elements participation and interaction with the educational learning process which leads to attract learners and increase their fortunate in learning. Facebook has become one of the most important informal cooperation tools between students as students began to use messages and groups on Facebook to communicate and coordination among them for business planning, collective projects, exchange opinion and advice in this regard.

Based on the high rate of using among college students that reach up to 85%, there are many unique features that make social media usable in educational activities. Such as instant messages, e-mail, post videos and pictures, and anyone can publish information and cooperation within an integrated system. More recently, Facebook

management developed a range of downloadable applications which could increase the educational means in Facebook while many of these technological tools reflect those found in educational programs used currently, as in the interactive whiteboard and module etc. Based on the above it can be said that creating new knowledge for graduate students may be affected in some way with massive successive developments that occurred in communities, especially with regard to social networks appearance such as Facebook and Twitter above all. Thus, the main objective of this study aims to detect the same in an academic environment which is King Abdul Aziz University, especially among graduate students using descriptive survey method through questionnaire which is the main study tool. The most important result that we achieved it shows that the graduate students use social media to get new knowledge with medium degree.

2 LITERATURE REVIEW

Through intellectual production of literature review, it had been reached to some of relevant studies with the current research, which is "The role of Social Media in creating new knowledge for graduate students at King Abdulaziz University in Jeddah". The **first study [3]** entitled "A Generic Framework for Extraction of Knowledge from Social Web Sources (Social Networking Websites) for an Online Recommendation System" aims to develop a framework to extract knowledge

from different social networks on the Internet for undergraduates in order to assist them in making critical decisions in their educational and cognitive lives. Knowledge is gathered from social networks on various Internet sources, regulated properly, and stored according to the current ontology in data warehouse, in addition to improving the quality and level of interaction between undergraduates on the Internet through social networks. From the study results; it appears that the article provide an appropriate proposal system to the students in terms of providing the student with appropriate professional direction and helping them through the proposed system. So that the students could get abundant, heterogeneous and unstructured knowledge efficiently. As knowledge is provided in various social networks with different forms, it needs to be structured and treated to make them accessible to everyone.

The second study [4] entitled “SOCIAL NETWORKING AS AN ALTERNATIVE ENVIRONMENT FOR EDUCATION” aims to analyze the impact of social networks on the educational process and knowledge in Romanian higher education and to clarify the value of the educational and cognitive social network sites for higher education. The study also aims to provide a typical proposal to implement Facebook usage in learning and knowledge transfer processes in higher education and find out the reason of university students joining to social networks. From the study results, we see that social network sites have become very popular among undergraduates and may considered valuable tools for education and knowledge transfer. The main advantages of using social network sites as a tool for education and knowledge transfer are the following: it is the unique approach which involves collaborative group work and knowledge share. Social network sites offer many benefits through the facilities provided for users of students such as facilitate communication, exchange knowledge and access to resources that can be discussed and easily accessible.

The third study [5] entitled “Constructive Disruptions for Effective Collaborative Learning: Navigating the Affordances of Social Media for Meaningful Engagement” aims to identify the social media used in the field of higher education in universities in South Africa through which knowledge could be transferred and shared so that; access to some constructive scientific knowledge gains for teachers and academics. This study also examines the existing interaction between the lecturer and student on the one hand, and between students and their colleagues on the other hand, in an environment of enhanced social media for knowledge transfer from proper resources. From

the study results, it is clear that social media technology grants an opportunity for detecting the repeated questions from undergraduates community, such questions with its responses represents a challenge for who are working on these media. Also, social media have a platform for diagnosis and assessment of these wrong cognitive concepts known by the student. Finally it is proved that social media enhances the ability of academics from faculty members to modify and adjust teaching methods, complex concepts and other cognitive issues. Also, the social media has created a suitable knowledge environment and interesting learning environment.

The fourth study [6] entitled “Collaborative Learning Experiences Using Social Networks” aims to search on the need of developing the cognitive and educational models rapidly in order to meet the speed of academic community development, also provide many technical solutions that elicited by information technology organizations’ leaders to facilitate knowledge transfer and collaborative learning using social media, in addition to providing some academic experiences on how to use social networks to improve the learning experience, transfer, acquire and store knowledge for students. The study concluded that social networks are used as collaborative learning environments in academic environments. Undergraduates prefer to learn how to use all new advantages in social networks. The social networks are useful as a tool for knowledge management. The students also prefer to use and dependent on such networks as a tool for managing, organizing and transfer knowledge. Finally the social networks represent a supportive useful tool to facilitate the process of education and encourage on cooperation and receive knowledge among students during learning.

3 PROBLEM STATEMENT

Since “Randy Conradz” [4] put the first pillar of social media sites whenever founded the first site to connect with his friends and colleagues named “com.classmates”; social media sites become the most important mean used by university students. Social media succeed as a distinct and useful interactive mediator at different levels of educational work on web since its establishment. Such media also have revolutionized and become big jump in communication world; as it allows the individual to communicate with his colleagues in all over the world. These media provide knowledge transfer and different experiences share. In the light of the foregoing; a desire in recognizing the role of social media in creating new knowledge among

university students has been emerged. Accordingly, our research questions are as follows:

- What is the role of social media in creating new knowledge for graduate students at King Abdulaziz University, Jeddah, Saudi Arabia?
- What is the degree of graduate students' use of social media at King Abdulaziz University?
- How do graduate students use social media in getting new knowledge?

4 METHODOLOGY OF SOLUTION AND METHODS OF DATA COLLECTION

This research follows the descriptive survey approach as it is more appropriate for social reality in order to understand the phenomena and draw features by collecting some data then analyze this relevant data, extract results and give recommendations.

- The study community and samples:
The study community consisted of all graduate students at King Abdulaziz University in Jeddah. A random sample was chosen consists of (64) student.
- The study tools:
In order to achieve this study purposes; the researcher used a questionnaire as it is suitable for this study purposes.

• Study limits:

1. Objective limits: this study is limited to this objective limits "The role of Social Media in creating new knowledge for graduate students at King Abdulaziz University in Jeddah".
2. Spatial limits: shall be limited to the graduate students' spatial limits at King Abdulaziz University?
3. Time limits: during the second semester of academic year 2015/2016

5 THE NUMERICAL RESULTS

This study aims to investigate the role of Social Media in creating knowledge among graduate students at King Abdulaziz University in Jeddah and how to use social media to get new knowledge. We will review, discuss and explain the field study results as follows: the female researchers found that graduate students use social media to get new knowledge with medium degree, as arithmetic mean was (3.34) and the standard deviation was (0.78).

From the above table1, we found that there are 10 uses with high degree. Arithmetic means ranging between (4.12–3.52), the reason may be related to the nature of the study and curriculum subjects for graduate studies required refer to databases and access to literature review in specific field.

Table 1. Arithmetic means and standard deviations to estimate the students usage degree of social media in descending order (Sample = 64).

No	Paragraph	Order	Arithmetic mean	Standard deviation	Usage degree
7	Log in information and data base	1	4.12	0.94	High
2	Get educational material for purposes of developing the acquired knowledge	2	4.08	1.17	High
15	Log in social networks such as Facebook through blogs related to any subject relevant to my study	3	3.98	0.98	High
17	Open chatting group with my female colleagues for educational purposes	4	3.97	1.14	High
19	Loading or uploading a file during my study for any subject	5	3.62	1.37	High
21	Using Youtube in studying the curriculum subjects	6	3.61	1.19	High
16	Organize the content of the interactive social network page in order to develop the acquired knowledge	7	3.57	1.32	High
14	Use applications of 2G technology in developing my knowledge in respect of any subject related to my specialize	8	3.54	1.29	High
8	For purposes of share pictures related to curriculums	9	3.53	1.28	High
9	For purposes of share videos related to curriculums	10	3.52	1.12	High
10	Designing educational blog related to specific subject	11	3.39	1.35	Medium
11	Designing educational wiki page in order to share knowledge regarding any subject related to my specialize	12	3.27	1.27	Medium
4	contact with professors for purposes of developing acquired scientific knowledge	13	3.26	1.29	Medium
3	Communication with colleagues in respect of curriculums	14	3.16	1.35	Medium

(Continued)

Table 1. (Continued)

No	Paragraph	Order	Arithmetic mean	Standard deviation	Usage degree
6	Benefit from curriculums provided on faculty members' sites or Facebook	15	3.15	1.11	Medium
1	Publish the educational material through the electronic link on social media	16	3.13	1.29	Medium
5	Publish my researches on social media in my blog	17	3.05	1.32	Medium
13	Link shares with other sites on social media in my blog	18	2.98	1.16	Medium
18	Use sites related to my specialization which depends on managing dialogue and discussions in this regard	19	2.98	1.19	Medium
20	Provide trusted e-learning resources such as electronic libraries sites for my colleagues in educational forums	20	2.83	1.22	Medium
22	Use interactive video in studying the curriculum subjects	21	2.77	1.11	Medium
12	Recall and browse my colleagues' shares in the blog through RSS	22	2.07	1.08	Low
Usage degree			3.34	0.78	Medium

Table 2. Arithmetic means and standard deviations in order to estimate how graduate students employed "usage" of social media in descending order (Sample = 64).

No	Paragraph	Order	Arithmetic mean	Standard deviation	Usage degree
25	Participate in developing the educational content through developing knowledge related to curriculum	1	4.10	0.93	High
37	Increase educational quantity	2	3.59	0.91	High
24	Benefit from writing reports	3	3.56	1.03	High
27	Solving problems related to the educational content	4	3.48	1.13	High
33	Informing students with all new	5	3.45	1.16	High
34	Better organization for the curriculum	6	3.35	1.22	Medium
35	Achieving the concept of active learning	7	3.17	1.24	Medium
43	Following up symposium and scientific meeting news	8	3.15	1.06	Medium
44	Get any documents for the purpose of research	9	3.10	1.08	Medium
32	Promoting the concept of continuing education among students	10	3.05	1.12	Medium
28	Encourage me on self-learning	11	2.95	1.15	Medium
29	Support cooperative and group learning	12	2.89	1.06	Medium
30	Reconsider individual differences among students	13	2.85	1.08	Medium
36	Organizing Chat groups	14	2.82	0.98	Medium
39	Improve communication skills of students	15	2.79	0.92	Medium
40	More democratic educational process	16	2.77	0.91	Medium
26	Provide me with feed back	17	2.65	1.10	Medium
23	Training on pass experimental tests	18	2.61	1.01	Medium
31	Benefit from interactive multimedia service	19	2.55	1.44	Low
38	Promoting the relationship between the student and professor	20	2.54	1.35	Low
41	Makes up for some office hours	21	2.53	1.36	Low
42	Following up the current events, activities and news related to the specialization	22	2.50	1.22	Low
"How to use" degrees			3.19	0.95	Medium

Diversity in the nature of specialization contributes refer to databases to get knowledge and because of the importance of 2G applications of blogs, wikis, interactive video and YouTube in acquiring knowledge related to graduate studies curriculum. For the need and importance of employing methods

of social media to create, obtained and exchange knowledge among female students at the stage of Graduate Studies. As the nature of curriculum study based on participation and cooperation through exchanging data and obtains ideas across social media that shall employ these methods.

Recall and browse my colleagues' shares in the blog through RSS phrase was Law because of many technical obstacles to many female students and the lack of skills that enable them to participate and interact in employing 2G of e-learning such as RSS technology Wikis and blogs. Female researchers found that the degree of employing graduate students "use" of social media to get knowledge with medium degree from their point of view as arithmetic mean was (3.19) and the standard deviation was (0.95).

From Table 2, we found that aspects of graduate students' uses employment of social media in getting knowledge with high degree in five aspects. Arithmetic means ranging between (4.10–3.45), the reason for this is that graduate studies female students exchanges files related to a specific subject across email or download sites which contributed in developing the curriculum content and enriches learning process. This was at the first level may be for the differences between female students abilities to acquire knowledge; therefore, they participate in data transfer to develop content that shall be directed to start self-education at this stage.

Regarding the low grade phrases in terms of employment for getting new knowledge; arithmetic means ranging between (2.50–2.55), the reason may be related to the lack of interactive multimedia in respect of the curriculum educational content of graduate studies, the lack of some female students skills in designing educational software serving the specialized study subject, its appearance contributes with low grade. The reason may be for the lack of electronic communication via social media between some professors and students which contributed to employ it with low grade to get knowledge. Meetings in office hours is more useful and effective in understanding the educational content which designed by the professor that is better than open resource which we cannot guarantee its accuracy or documentation.

6 CONCLUSION AND RECOMMENDATIONS

The presented field study was done about the role of social media in creating new knowledge for graduate students at the University of King Abdulaziz in Jeddah (Saudi Arabia) revealed many results that highlight the degree of use social media and the most prominent kind in use. We conclude from the foregoing that graduate students at King Abdulaziz University in Jeddah are getting help too much from social media in many purposes especially Facebook, YouTube and 2G technologies application as those media providing tools that contribute in developing their knowledge and enriches the same in topics

related to their different specialization through which students could link and support their materials and curriculums with countless videos and pictures on social media as such media contains many blogs related to specialists in the field which contributes in getting needed knowledge from trusted resources in their scientific field.

We cannot also lose sight of a very important factor for the graduate students' tendency to use social media to increase and develop their knowledge because knowledge and events are constantly renewed in these media; therefore they will have the latest trends in their specialization. Through social media; they could open chat groups, so that communicate with each other will be easier in the event of group research, joint studies and workshops.

We recommend training should be provided to graduate students more and more on the employment of technologies and social media in creating knowledge and obtain it as results revealed that the degree of use was medium. Working on developing graduate students skills and techniques of using database, training them on using sites that rely on managing dialogue and discussions, use interactive programs to exchange posts with female students, acquire skill of attaching (files) with e-mails and send them to female colleagues, as results revealed that its use is medium. We also recommend the need for directing professors to the importance of communicating via social media with graduate students, as the effective and continuous communication have the greatest impact on disseminating and creating knowledge as well as the greatest impact on maintaining and developing the concept of continuing education.

REFERENCES

1. Anzy, C. "Effectiveness of using social networks in getting sciences and the trend towards knowledge society for third-grade female students in Medina". Unpublished PhD thesis, University of Umm Al-Qura: Mecca 2013.
2. Abu Sailik, Z. "The impact of electronic social networks on undergraduates at Jordan universities and its proposed role in developing their balanced character" Unpublished PhD thesis, University of Jordan: Oman, 2012.
3. Javubar Sathick, Jaya Venkat.—A Generic Framework for Extraction of Knowledge from Social Web Sources (Social Networking Websites) for an Online Recommendation System.—Vol. 16, No 2, 2015.—Available At: <http://www.irrodl.org/index.php/irrodl/article/view/2093/3275> (1 Dec, 2015).
4. Andrei Stanciu, others.—SOCIAL NETWORKING AS AN ALTERNATIVE ENVIRONMENT FOR EDUCATION, 2012.—Available At: ftp://ftp.repec.org/opt/ReDIF/RePEc/ami/articles/11_1_4.pdf (1 Dec, 2015).

5. Patient Rambe. "Constructive Disruptions for Effective Collaborative Learning: Navigating the Affordances of Social Media for Meaningful Engagement".—Electronic Journal of e-Learning.—Volume 10, Issue 1, 2012.— Available At: <file:///C:/Users/pc1/Downloads/ejel-volume10-issue1-article184.pdf> (1 Dec, 2015).
6. Arturo Mora, others.—Collaborative Learning Experiences Using Social Networks.—2009, available At: http://www.academia.edu/194162/Collaborative_Learning_Experiences_Using_Social_Networks (1 Dec, 2015).
7. El Askary, Aboud Abdullah, "The research methodology in humanities" 2004 ed., Al Nemr publication house: Damascus, Syria.
8. Salim, Ahmed Jamal: "Social media sites, advantages, disadvantages and what is the correct usage", available At: <http://www.alukah.net/culture/0/63253/#ixzz3p1wWhNo> (2 January, 2016).

Fuzzy decision trees for text document clustering

Wahiba Ben Abdesslem

Department of Computer and Information Technology, University of Taif, Taif, Saudi Arabia

Kawther Dridi

The High Institute of Management of Tunis, Tunis, Tunisia

Eman Alkhamash

Department of Computer and Information Technology, University of Taif, Taif, Saudi Arabia

ABSTRACT: Due to the large number of documents available on the web, operations such documents' retrieval become difficult task. This has led to the development of a variety of classification algorithms to resolve this kind of problems. In classification problems, we are often confronted with overlapping classes, which have limitless boundaries, and cannot be presented in a specific partition. In this paper, we propose an approach based on fuzzy logic and distance measures for text document clustering. The key idea is to search for the similarity and dissimilarity between documents. An experimental study using biomedical area articles extracted from Jabref is conducted. The experimental results are analyzed and discussed and found to be promising for text document clustering.

1 INTRODUCTION

Clustering is an important task for assigning objects into clusters or groups where collection of objects of the same class are similar and objects from different class are dissimilar. Examples of some famous clustering tools are the C4.5 algorithm (Quinlan, 1993), k-mean clustering algorithm (Chen et al., 1998) and Self-Organizing Map (SOM) (Kohonen et al., 2001), (Vesanto & Alhoniemi, 2000).

In the logic of classic clustering, each object has variables with accurate and precise values. Thus, each object belongs to a single class. However, information is not always accurate and complete and can be uncertain and imprecise.

For example, if a person says that the book he has bought is not too expensive, this is relative and depends on the perception of individuals. For another person, whose resources are limited, the book may seem to him too expensive. Individuals can find themselves in different situations, where they use incomplete information, and need to make decision according this information. Fuzzy logic is the solution to this problem and allow to handle this kind of information (Zadeh, 1965).

Due to the large number of documents available on the web, operations such as: documents retrieval, document classification, document clustering, become difficult tasks. In classification, for example, we are often confronted with overlapping classes, which have not limited boundaries and cannot belong to a specific class.

So to help preventing this issue, we propose a new approach that combines fuzzy decision tree and document similarity for document clustering. The key idea is to search for the similarity and the dissimilarity between documents to facilitate classification. Our approach consists of three stages. In the first stage, we collect a set of documents. In the second stage, documents are cleaned by decomposing them into words, unnecessary words are eliminated and represented in a formal representation using vectors. We also use the Cosine similarity distance to calculate the similarity between document vectors. Finally, we perform documents clustering, using a proposed fuzzy clustering algorithm.

The remaining sections of this paper are organized as follows. The background of fuzzy logic and fuzzy decision trees, are presented in section 2. In section 3, we provide a description of the proposed fuzzy decision trees algorithm for document clustering. An experimental study, to evaluate the usefulness of fuzzy algorithm in text document clustering, is presented in section 4. Summary of concluding remarks is included in section 5.

2 BACKGROUND

2.1 Fuzzy logic

Fuzzy logic is conceived as an extension of classical logic. It is introduced by Lotfi Zadeh as a mathematical model characterized by its ability to

represent incomplete and imprecise information (Zadeh, 1965). Fuzzy logic uses probability theory and other theories such as Demster-shafer evidence theory, possibility theory and so on (Zadeh 1965).

Fuzzy logic describes problems as If-Then rules expressed in natural language. Unlike Boolean sets, where the characteristic function takes only two values either 0 or 1, the function of the fuzzy logic can take current values between 0 and 1. Therefore fuzzy logic considers the notion of belonging of an element to a set of classes as a function that can take values from 0 to 1. In other words the difference between the classical sets and the fuzzy sets is presented by introducing a membership function.

To understand fuzzy logic concept, assume that we have fuzzy set E that is characterized by a membership function μ_E . μ_E associates to each object in the universe, its membership degree $\mu_E(x)$, in the interval [0, 1]. Assume that X is a collection of objects. X is called the universe of discourse and x represents an element of X. More $\mu_E(x)$ tends to 1, more x belongs to E. A fuzzy set is constituted by a set of fuzzy subset.

For example, let's consider three fuzzy sets such as "young", "mature" and "old" which will be characterized by a membership function μ_E associated to each person in the universe X. X is a collection of persons $X = \{x_1, x_2, x_3, \dots, x_n\}$ where the degree of membership function is expressed by a real number in the interval [0, 1]. So this degree can be interpreted as a continuous process in which the membership of a person to the set of adults takes values between 0 and 1 as shown in Figure 1.

Figure 1 shows three fuzzy sets "young", "mature" and "old" with three examples of membership function, related to age, in the interval 0 to 70 years. The degree of membership takes value from 0 to 1. The three functions mentioned the membership of a person in the sets of young, mature and old ages. So, we can interpret this degree as follows: If a person is a 20 years old for example, his degree of membership in the set of young persons is 1, in the set of mature is 0.35 and in the set of old persons is 0. If a person is a 30 years old, his degree of membership in the set of

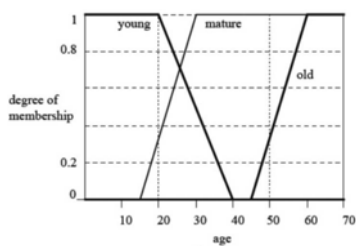


Figure 1. Example of membership function.

young persons is 0.45, 0.9 in the set of mature but in the set of old persons. Another example when a person is a 50 years old, the degrees of membership are 0 in the young set, 1 in the mature and 0.3 in the set of old (Zadeh, 1965).

Over the past years, fuzzy logic has gained considerable attention in different fields for example machine learning, data mining field and many other fields (Hüllermeier, 2005). For instance, one of the most applications in the fuzzy logic area is Sendai Subway system in Sendai, Japan, was recently addressed using fuzzy logic (Bart & Satoru 1993). The control of the Nanboku line, developed by Hitachi, employed a fuzzy controller to direct the train. Researchers used fuzzy rules to accelerate, slow and brake the subway trains more smoothly (Bart & Satoru 1993). Moreover, fuzzy logic has been also used for air conditioning system; These systems use fuzzy logic to control the heating and cooling, this saves energy by make system more efficient. Therefore fuzzy logic represents the notion of belonging of an element to a set of classes as a function that can take values from 0 to 1.

2.1.1 Fuzzy operators

In the case of traditional logic, the operators used are AND, OR and NOT to express the intersection, union and the complement operations but in fuzzy logic these operators need to be defined as functions for all possible fuzzy values from 0 to 1 (Zadeh, 1965), (Zimmermann, 2011). So a generalized form of these operators will be useful. The generalization of these three operators are presented in Table 1 where A and B are fuzzy sets and x is an element (Zadeh, 1965), (Zimmermann, 2011).

The Union, Intersection, and Complement are defined in (Zadeh, 1965), (Zimmermann, 2011) as follows:

Union: A fuzzy set is completely determined by its membership function. The union of bivalent fuzzy sets with membership functions is defined as the maximum. The membership for the union of two sets A and B is:

$$\mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x)) \quad \forall x \in X \quad (1)$$

Functions used to interpret union are denoted as T-conorms. The union operation is the equivalent of the OR operation.

Table 1. Fuzzy operators.

A and B	Min (A,B)
A or B	Max (A,B)
Not A	1-A

Intersection: The intersection of bivalent fuzzy sets with membership functions is defined as the minimum. The membership for the intersection of two sets A and B can be defined as follows:

$$\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x)) \quad \forall x \in X \quad (2)$$

Functions used to interpret intersection are denoted as T-norms. The intersection operation is the equivalent of the AND operation.

Complement: The complement of fuzzy set A with membership function is defined as the negation. The complement of membership function can be defined as follows:

$$\mu_{A^c}(x) = 1 - \mu_A(x) \quad \forall x \in X \quad (3)$$

Functions used to interpret complement are denoted as S-norms. The complement operation is the equivalent of the NOT operation.

2.1.2 Membership function

The membership function is used to measure the membership degree of every element to its corresponding set. It associates each element to its degree.

For Boolean sets, membership degree is either 0 or 1 but in the fuzzy case membership degree is expressed in the interval [0, 1].

2.2 Fuzzy decision tree

In practical classification problems, one is often confronted with overlapping classes, whose own borders enclosed, and cannot be defined by a clear partition.

More specifically, in the case of traditional decision tree, the boundaries used for defining predicates at inner nodes have been criticized. So to resolve the decision boundaries, an obvious idea is to make fuzzy predicates at the inner nodes of a decision tree. This is the main motivation underlying fuzzy extensions to decision tree.

Fuzzy decision tree for document clustering is based on the concept of fuzzy sets. It considers inaccurate information, and clusters with wavy borders. It will generate clusters with indistinct borders. For example, a document can belong to multiple clusters with a certain degree.

Fuzzy decision tree offers an extension to traditional decision tree. It was recently presented as a novel method for classification (Bouchon-Meunier et al., 1993), (Bouchon-Meunier & Marsala, 1999), (Wahiba & Ahmed, 2016).

Roughly speaking it is considered as a kind of generalized description of class. In fuzzy clustering an object may be grouped in different clusters at the same time with a particular degree. The degree

with which it belongs to a particular class is defined in terms of a membership degree.

Over the past years, fuzzy decision tree have attracted attention in different fields for example recognition of graphic symbol and *text documents clustering and visualization* (Dash et al., 2012). To cluster sets of textual documents based on their similarity, Chang et al. (Chang et al., 2014) created an evolving tree which is a tree that contains root nodes, trunk nodes and leaf nodes when the root node is the first created node in the tree, trunk node is connecting the leaf nodes and leaf nodes are the formed clusters (Chang et al., 2014).

3 PROPOSED APPROACH FOR DOCUMENT CLUSTERING

The approach consists of three main steps: document collection, document processing, and document clustering as shown in Figure 2.

In the first step, we collect a set of documents in different domains such as scientific documents, historic documents, mathematical documents etc. The second step is processing the extracted documents. In other terms the processing task consists on text cleaning and representation such as vector representation, textual representation, *TF IDF* representation etc. The cleaning task consists of decomposing document into separate words and eliminating the unnecessary ones. The representation task is characterized by modeling the document as a vector.

In the third step, we have two sub steps: (i) we calculate the terms associated to its frequency in the text file for example the words “fuzzy” appears just one time, but the words “classification” appears 3 times in the abstract. Calculation results will be represented into a matrix to facilitate the comparison between documents. In fact, documents will be compared by the number of terms that appear in the text. (ii) Each document will be associated to the specific cluster (the nearest neighbor cluster).

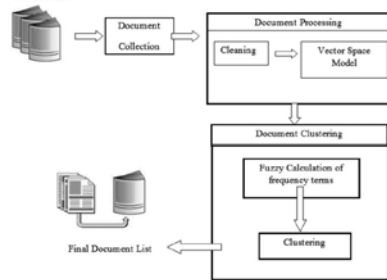


Figure 2. Document clustering architecture.

In the next subsections we describe each step in details.

3.1 Collecting documents

The documents collection represents the first step of this approach. To do so, we make use of Jabref (Jabref, 2016). It is free software that returns a list of relevant documents from a previously chosen database. It is also, a tool for importing data from online scientific databases like MEDLINE.

3.2 Processing documents

The selection step consists of the extraction of pertinent abstracts from MEDLINE database. MEDLINE articles are accessible through the PubMed server (Medline, 2016). At this step we must guarantee that all abstracts are randomly selected and without any user intervention. After finishing the selection step, the output must be cleaned by removing all the useless information that affects the results performance. The goal of this step is to optimize the performance of the next step.

Our selected documents are presented by two tasks the cleaning and the vector representation task.

3.2.1 Cleaning documents

The cleaning task is characterized by deleting the unnecessary words, also known as a stop words. This phase consists of decomposing document into separate words and eliminating the stop words. In addition, the cleaning task is used to reduce the textual data size and improve efficiency and effectiveness.

An extra of English stop words is given by the figure 3 as follow:

To understand the cleaning step role we present an example of the cleaning of an abstract extract given by the table 2. This table is composed by two columns the first one represents the abstracts from the text files forming the document collection but the second represents the abstract extract after the cleaning process.

3.2.2 Vector representation of documents

Vector model is a transformation of document from textual form to an algebraic one. In this work we use the vector representation thanks to its ability to facilitate the calculation of terms. It is a mathematical model that represents documents as vectors. These vectors contain the document terms associated to their weights. So to calculate these weights we use the bag of words model, one of the most weighing technique.

In others words, a vector contains a word from the input text associated to its frequency in the text file. An example of a text vector representation is presented by the table 3. This table shows the

Stopwords		
a	it	these
about	its	they
again	itself	this
all	just	those
almost	kg	through
also	km	thus
although	made	to
always	mainly	upon
among	make	use
an	may	used
and	mg	using
another	might	various
any	ml	very
are	mm	was
as	most	we
at	mostly	were

Figure 3. Stop words.

results of vector representation process. The first column contains the text extract after the cleaning step but the second contains the vector representing the input text. For example the word “plasma” appears just one time in the abstract but the word “biomarkers” appears two times in the abstract.

Formally, the vector is given by the following formula:

$$V_j = (\Pi_{d_j}(t_1), \Pi_{d_j}(t_2), \Pi_{d_j}(t_3), \dots, \Pi_{d_j}(t_i))$$

Where

V_j : vector that represent the document j

D_j : document j

T_i : term i

$\Pi_{d_j}(t_i)$: The membership degree of the term i in the document j

This formula is composed essentially by $\Pi_{d_j} t$. $\Pi_{d_j} t$ represents the membership degree of the term t in the document d. It is obtained by measuring the weight or the number of occurrence of the term t in the document d.

The representation step consists of the representation of each document by a vector whose components are the words contained in the text. A collection of text documents can be represented by a matrix where rows are the words (Terms) that appear at least once and columns represents clusters as shown in the Table 4:

3.3 Clustering documents:

We choose to cluster the text document with fuzzy clustering algorithm. The clustering with fuzzy clustering algorithm is unsupervised classification and provides a clustering without a priori known number of classes that’s means we choose an arbitrary k clusters.

This type of method assumes that documents which belong to a same class have similar characteristics. The main idea is to define a center for

Table 2. Example of text cleaning.

Input text	Output text
Recently some plasma biomarkers of inflammation have been recognized as important cardiovascular risk factors There is little information about the effects of aerobic exercise training on these biomarkers and the risk of metabolic complications in obese type 2 diabetes patients	Recently some plasma biomarkers inflammation have been recognized important cardiovascular risk factors little information effects aerobic exercise training biomarkers risk metabolic complications obese type diabetes patients

Table 3. Example of text vector representation.

Input text	Output text
Recently some plasma biomarkers inflammation have been recognized important cardiovascular risk factors little information effects aerobic exercise training biomarkers risk metabolic complications obese type diabetes patients.	(Recently, 1), (some, 1), (plasma, 1), (biomarkers, 2), (inflammation, 1), (have been, 1), (recognized, 1), (important, 1), (cardiovascular, 1), (risk, 2), (factors, 1), (little, 1), (information, 1), (effects, 1), (aerobic, 6), (exercise, 6), (training, 6), (metabolic, 2), (complications, 1), (obese, 4), (type, 4), (diabetes, 4), (patients, 4)

each class. Each class is characterized by its center noted C_i and the number of elements noted N_i .

The classification is carried according to the following algorithm:

- i. Choose an arbitrary k clusters.
- ii. Calculate cluster centers.
- iii. Calculate membership degrees of terms in documents.
- iv. Repeat (ii) and (iii) until emergence is stable

Each step is explained is follows:

- i. K can vary from 3 to 5, 7..., it should be an odd number.
- ii. We calculate the centers vectors by the following formula:

$$C_j = \frac{\sum_{i=1}^N \prod C_j(ti) * ti}{\sum_{i=1}^N \prod C_j(ti)} \quad (5)$$

Where:

C_j : the center

$\prod_{ij}(ti)$: Membership degree of the term i in the cluster j.

- iii. A similarity measures between two documents: d_{i+1} and d_i is computed.

In this study, the cosine similarity between d_{i+1} and d_i , is computed as follows:

$$\cos(\theta) = \frac{X.Y}{\|X\| \|Y\|} \quad (6)$$

We used the Cosine similarity distance to calculate the similarity and the dissimilarity between

Table 4. Membership matrix.

	T1	T2	T...	Tn
C1				
C2		$\prod_{D2}(t2)$		
C...				
Cn				

vectors. In our work this metrics represents the distance separating the document i from the document j.

We decided to use the Cosine similarity metric than the distance Euclidean metric because the first one focuses on the presence and the absence of terms in documents.

For example we will consider the following vocabulary with three documents to prove the use of Cosine similarity metric.

Vocabulary: < patient, diabetic, system, cardiovascular>

Document1 is presented as follow:

Vector1: < 1, 1, 1, 0 >

Document2 is presented as follow:

Vector2: < 0, 0, 1, 0 >

Document3 is presented as follow:

Vector3: < 0, 0, 0, 0 >

Where

1: means the membership of the term in the document.

0: means the absence of the term in the document.

To calculate the similarity between documents we can use different metrics such as Euclidean distance, Cosine similarity distance. We opted for cosine similarity measure because, we have noticed that the cosine similarity measure gives more better results than other similarity measures such as the Euclidean distance.

iv. Repeat (ii) and (iii) until emergence is stable ($\|U(k+1) - U(k)\| < \epsilon$). Finally, the documents are clustered and a final list of documents is obtained.

4 EXPERIMENTATION AND TEST

A case study to evaluate the effectiveness of the fuzzy proposed algorithm as a text document clustering tool was conducted. In this study the abstracts of 100 randomly selected articles from Jabref were used for experimentation to observe the classification process. A predefined list of stop words, which consisted of 238 words, was used. The parameters used were: $k = 3$ clusters. In this study a laptop with an Intel core i2 processor, and Windows (64-bit) was used. All our experiments were implemented using java language and compiled in Eclipse framework.

The experimental results in terms of classification rates and computational complexity are analyzed and discussed.

To validate our model we need to evaluate its implementation and we also need to test it on a set of adequate data. For this reason we decided to use two data sets.

First of all, we are going to collect a set of documents in different domains. This step is named the selection step. It is the first step of our work. The selection step consists of the extraction of 100 abstracts from MEDLINE data base.

To measure the efficacy of our proposed approach, we propose to use one of the most popular measure classification rate which is the accuracy rate (Sokolova & Lapalme, 2009). It is measuring the Overall effectiveness of a classifier system. The formula is as follows:

$$\text{Classification rate} = \frac{\text{the number of documents correctly class}}{\text{the total number of samples in the test}} \quad (7)$$

In this experiment we used the data extracted from the Jabref system. For each data set we calculate the classification rate using both fuzzy classification algorithm and C4.5 algorithm. Table 5 presents the classification rates of different systems:

Table 5. Classification rates of different systems.

Data set	Algorithm	Accuracy
Fuzzy logic documents	C4.5	96,1%
	Fuzzy classification algorithm	97%
Biomedical documents	C4.5	73,1%
	Fuzzy classification algorithm	75%

Table 5 describes our test results for the two data sets fuzzy logic Medline and biomedical Medline.

The results are encouraging since the classification rates obtained are comparable.

On one side, the C4.5 showed very good results thanks to its ability to classify all the documents for the two datasets, but it was unable to detect the similarity between them.

On the other side, the fuzzy classification algorithm showed its efficiency to classify all the documents, but also it has an important characteristic over the C4.5 by its ability to detect similar documents.

In figure 4 we used the first data set fuzzy logic. To draw the curve blue, we began by testing or evaluating our algorithm for only 10 documents and we recorded the corresponding run time. Then, we incremented our tested documents by 10 each time until completing our 100 total documents. We followed these same steps to draw the red curve using the C4.5. In Figure 5, we used the second Medline (Biomedical). To draw the corresponding curves, we followed the same steps as the first Medline.

We noticed according to figure 4 and figure 5, that the proposed algorithm takes less times running than the C4.5 regardless of the size of the data set.

We tested, also, our approach using different number of clusters, showing the tree size and the accuracy rates as shown in Table 6.

Table 6 summarizes the number of clusters created, tree size and the classification rates. The results shows that the classification rates decreased as the number of clusters increased and the tree size increased as the number of clusters increased.

Figure 6 summarizes the time required for the classification process for the number of clusters = 3, 5, 15 respectively. The time axis shows the time in seconds required to classify a new abstract. Moreover the Figure 6 shows that the time required for the classification process increased as the number of abstracts increased.

With the proposed approach, documents imported from Jabref tool could be clustered and visualized as a list of text documents. It is expected

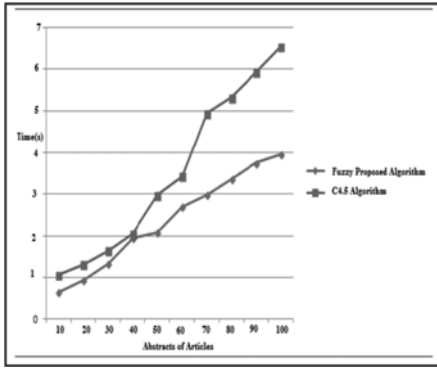


Figure 4. Run time curve for Fuzzy logic medline.

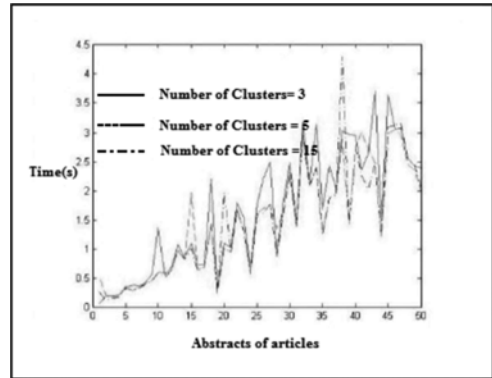


Figure 6. The learning time with 3, 5, 15 clusters.

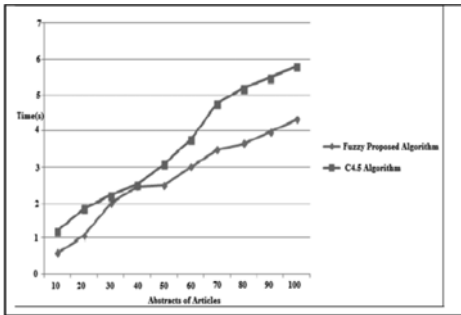


Figure 5. Run time curve for biomedical medline.

Table 6. The proposed algorithm with 3,5,15 clusters.

Algorithm	Number of clusters	Tree size	Accuracy
Fuzzy proposed algorithm	3	5	97%
Fuzzy proposed algorithm	5	7	96%
Fuzzy proposed algorithm	15	27	75%

that the computational complexity increased with increasing the number of documents used for clustering.

5 CONCLUSION

Nowadays getting the right information at the right time has become a necessity. However upon classification of text document, a classifier may

be not able to determine the class associated with a document. Therefore our need for an application to organize and to classify text documents is becoming more and more crucial.

In this paper, we proposed a new approach to classify documents based on fuzzy decision tree. This approach begin first, by collecting a set of documents. Then, these documents are cleaned and represented in a formal representation: the documents are represented as vectors. Finally, the Cosine similarity distance is used to compute the similarity between document vectors and perform documents clustering.

In an experimental study, we used two datasets grouping documents related to fuzzy logic and biomedical Medline. Then we compared the fuzzy classification algorithm with the C4.5 algorithm by measuring the classification rates and the running times.

The results we obtained are promising. They showed that fuzzy classification algorithm is competitive. They prove its efficacy not only to classify documents but also to detect their similarities. Moreover the results obtained shows that the complexity in terms of time increased with increasing the number of articles used for clustering and also with increasing the number of clusters.

REFERENCES

- Bart, K., & Satoru, I. (1993). Fuzzy Logic ||, retrieved from <http://Fortunecity.com/emachines/e11/86/fuzzy-log.html>.
- Bouchon-Meunier, B., M. Ramdani, & L. Valverde (1993). Fuzzy logic, inductive and analogical reasoning. In A.L. Ralescu (Ed.), Fuzzy Logic in Artificial Intelligence, Volume 847 of Lecture Notes in Computer Science, pp. 38–50. Springer.

- Bouchon-Meunier, B., & Marsala, C. (1999). Learning fuzzy decision rules. In *Fuzzy sets in approximate reasoning and information systems* (pp. 279–304). Springer US.
- Chang, W.L & Tay, K.M & Lim, C.P. (2014). *Soft computing in industrial applications*. Springer International Publishing.
- Chen, C.W., J. Luo, & Parker K.J. (1998). Image segmentation via adaptive k-mean clustering and knowledge-based morphological operations with biomedical applications. *IEEE Trans. Image Processing* 7(12), 1673–1683.
- Dash, S.K., Mohanty, G., & Mohanty, A. (2012). Intelligent air conditioning system using fuzzy logic. *International Journal of Scientific & Engineering Research*, 3(12), 1–6.
- Huang, L., Milne D., Frank E., & Witten I.H. (2012, August). Learning a concept-based document similarity measure. *J. Am. Soc. Inf. Sci. Technol.* 63(8), 1593–1608.
- Hüllermeier, E. (2005). Fuzzy methods in machine learning and data mining: Status and prospects. *Fuzzy Sets and Systems* 156(3), 387–406.
- Jabref (2016). <http://jabref.sourceforge.net/> /accessed, April 2016.
- Kohonen, T., M.R. Schroeder, & T.S. Huang (Eds.) (2001). *Self—Organizing Maps* (3rd ed.). Secaucus, NJ, USA: Springer—Verlag New York, Inc.
- MEDLINE.2016.<http://www.ncbi.nlm.nih.gov/pubmed/> accessed, April, 2016
- Quinlan, J. (1993). *Ross: C4. 5 Programs for Machine Learning*. San Francisco: Morgan Kaufmann Publishers.
- Sokolova, M. & Lapalme G. (2009, Jul). A systematic analysis of performance measures for classification tasks. *Information Processing and Management* 45(4), 427–437.
- Vesanto, J., & Alhoniemi, E. (2000). Clustering of the self-organizing map. *Neural Networks, IEEE Transactions on*, 11(3), 586–600.
- Wahiba, B.A., & Ahmed, B.E.F. (2016). New Fuzzy Decision Tree Model for Text Classification. In *The 1st International Conference on Advanced Intelligent System and Informatics (AISII2015)*, November 28–30, 2015, BeniSuef, Egypt (pp. 309–320). Springer International Publishing.
- Zadeh, L.A. (1965). Fuzzy sets. *Information and control*, 8(3), 338–353.
- Zimmermann, H.J. (2011). *Fuzzy set theory—and its applications*. Springer Science & Business Media.

Generation of use case UML diagram from user requirement specifications

Wahiba Ben Abdesslem & Eman Alkhammash

Department of Computer and Information Technology, University of Taif, Taif, Saudi Arabia

ABSTRACT: Modeling is an important stage during the software development, especially in analysis and design. The process of constructing models such as UML diagrams from user requirements specification is a difficult task as it requires manual analysis of the requirements document to generate the UML diagrams. Model transformations are needed to generate UML diagrams from natural language requirements. Tools that support such transformation, however, are still very limited. In this paper, we propose an approach that allows to generate a use case diagram from user requirements using NLP techniques.

1 INTRODUCTION

The success of software development is based heavily on the ability to develop programs, which respond to the needs expressed by its users. The task of transforming user requirements to UML models is costly and time consuming for designers. Finding automatic techniques that support such transformation would be of great benefit. In this paper, we present an approach that attempts to facilitate the process of mapping user requirements into UML use case diagram.

Based on the MDA approach (Figure 1), user's requirements are considered as a source model represented in natural language text. The source meta-model represents a description of the concepts of natural language texts (verbs, nouns,...). UML use case diagram is considered as the target model. The target meta-model is an XML file grouping the concepts related to use case diagram. Consequently, the approach consists of four steps:

- i. Defining a use case diagram metamodel, describing use case diagram, and requirement metamodel, describing requirements.
- ii. Defining transformation function, ensuring the transformation of the requirement metamodel into use case diagram metamodel. This function gathers a number of transformation rules. The developed transformation rules are added to Gate components: JAPE (JAVA Annotation Patterns Engine) and Gazetteer lists (Cunningham, et al., 2002).
- iii. Processing the text related to user requirements, using Gate tools (sentence splitter, tokenizer, POS Tagger), the result is a processed text with grammatical categories (i.e. verbs, nouns, adverbs, etc.).

- iv. Executing the transformation rules, using as input the results of the previous step, to generate an XML file. Finally, the XML file is transformed into a use case diagrams.

The transformation rules developed in the second stage are based on the two metamodels: the use case diagram metamodel as shown in Figure 2 and the user requirement metamodel as shown in Figure 3.

Figure 2 describes the concepts of use case diagram where a system is composed by uses cases and actors. The filled diamond denotes the composition relation.

The relationship between the composite (i.e. system) and the components (i.e. use case and actor) is a strong relationship. If the composite is destroyed, all the component parts must be destroyed. An actor is linked to use cases with association. Therefore, each association is created

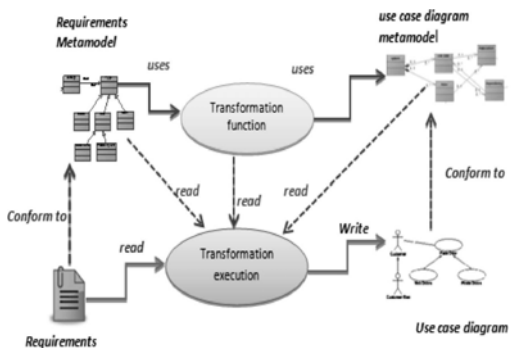


Figure 1. URS to use case diagram transformation framework.

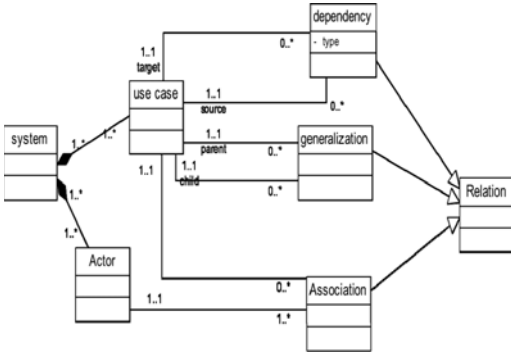


Figure 2. Use case diagram metamodel.

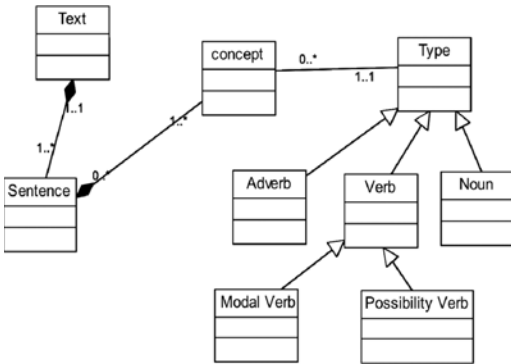


Figure 3. User requirements specification metamodel.

for one use case and one actor. Also the use cases can have relationships with each other's. The relations can be a dependency or a generalization. In the dependency relation, we indicate that a use case may be a source or a target of the dependency. The type attribute refers to include or extend dependencies. In generalization, we make use of parent and child to describe generalization between use cases.

Figure 3 describes the concepts of a natural language text: a text is composed by sentences. Each sentence is composed by terms or concepts, and have a specific types: adverb, verb or noun. The modal verb (describing ability, permission, requests and advice, such as: can, could) and the possibility verbs (e.g. may, might) are considered as subclasses of the class verb.

The rest of the paper is structured into the following sections: Section 2 describes the background. Section 3 describes the related work. Section 4 gives details about the proposed transformation approach. Section 5 is an evaluation of our proposed approach. Finally, the paper is concluded to discuss the future work.

2 BACKGROUND

2.1 Gate

Gate is text processing platform for language engineering developed at the University of Sheffield (Cunningham, et al., 2002).

Gate includes several plugins and other components that allow both the annotation and information extraction. Each annotation has start and end offset and set of features, each of which has a name and value. JAPE is part of the Gate system that execute JAPE grammar phases.

Each phase consists of set of action rules that consists of two parts. The left hand side part represents an annotation pattern whereas the right hand side part represent the action to be taken when the left hand side part is matched. Annotations created by different tools for language processing such as Gazatters. In Gate, Gazatters are used to find the occurrence of proper names and keywords in text.

2.2 Model-Driven Architecture (MDA)

MDA provides a frame work that uses models for software development. The Object Management Group (OMG) introduced MDA in 2001 (OMG 2016). Model transformation is one of the prominent features of MDA (Segura, et al., 2007), Model transformation is the process of converting one model to another within the same system (Dube & Dixit, 2012). The transformation function use transformation rules, to transform automatically an instance of the source metamodel, to an instance of the destination metamodel. The approach proposed in this paper uses transformation rules to transform user requirements specification to use case diagram.

2.3 UML use case diagram

Use case diagram is a UML diagram used to capture functional requirements and model the dynamic aspects of the system. Use cases play important roles in the early stages of the development of the system. A use case diagram represents an abstract view of the systems and isolates details to better understand the portion of the system of concern (UML, 2016). UML use case diagrams consist of four main elements which are: system, actors, use cases, and relations. The actors, depicted by stick man icon, represent system/people that interact with the modeled system. The use cases, depicted by named ellipses represent main functionalities provided by the system). The relations (association, dependency, and generalization) are used to indicate interactions between components. Association relationship is used between actors and uses cases. Generalization can be used between actors,

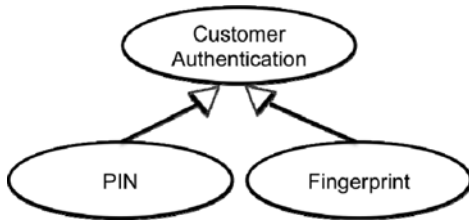


Figure 4. Generalization use case relationship.

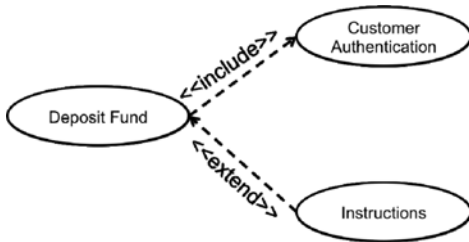


Figure 5. Use case relationships.

when they have the same roles. The commonly relationships used between use cases are: generalization and dependencies. Generalization is used to show a parent-child relationship between use cases; thus it used when two or more use cases have similar behaviors. Generalization is shown as a directed arrow with a triangle arrowhead. The child use case is connected at the base of the arrow as shown in Figure 4. The two use cases PIN and Fingerprint are childs of Customer Authentication use case.

There are two commonly relationships of dependencies in use case diagram, which are <<include>> and <<extend>>. The <<include>> is used when the base use case is incomplete without the included use case. For instance, Deposit Fund use case include Customer Authentication use case as shown in Figure 5. whereas the <<extend>> is used when the use case is independent (optional) on the extended (base) use case. For instance, Instructions use case extends Customer Authentication use case as shown in Figure 5.

3 RELATED WORK

In this section some works will be described related to automated generation of UML diagrams from user requirements.

Examples of such work are NL-OOPS (Mich & Garigliano, 2002), CM-Builder (Harmain & Gaizauskas, 2003), R-Tool (Afreeen, et al., 2011), (Vinay, et al., 2009), etc. NL-OOPS and CM-BUILDER use natural language specifications

to construct an object oriented analysis models, but the output is only a preliminary model which necessitates to be completed. Segura et al. (Segura, et al., 2007), propose the graph transformations which used to automate the transformations from Web requirement models to Web design models. However, this approach does not work with standard notations such as UML (Vinay, et al., 2009) describe a Natural Language based tool: R-TOOL, which aims at supporting the analysis stage of software development. R-TOOL analyses software requirements texts to generate use case, and class diagrams. The tool demonstrates just an initial experimental work, which has to be improved. In this paper, a new approach will be presented to map, automatically, user requirements into UML use case diagram. The proposed approach consists of, mainly, the following steps: first, the requirements and use case metamodels are defined. Second, transformation rules, allowing the transformation of the requirements metamodel into the use case metamodel, are constructed. Third, a text representing requirements is processed to identify its concepts. Then, transformation rules are executed, and finally, a use case diagram is represented graphically.

4 THE PROPOSED APPROACH

Transforming a model into another means, that a source model is transformed into a target model based on some transformation rules. The basic contribution of this work is in transforming the requirement (source model), into a use case diagram (target model). As known in MDA, the challenge is how to find transformations between two meta models, that can be applied to transform models. Different methods can be used for defining the transformation rules. The key idea is the relationship identification between the two metamodels. Since the two metamodels are defined (figure 2 and figure3), the rules are created based on the information in the source and target metamodels.

The following sub-sections describe in details the proposed approach.

4.1 Transformation rules construction

UML use case diagrams consist of four main elements which are system, actors, use cases, and relations (association, dependency, and generalization). Specific rules will be identified, to extract each concept from the requirements.

4.1.1 Extracting system concepts

The system indicates the scope of the system. In requirement specifications, the system is mapped


```

Phase : System
Input : LookupToken
Options : control = applet
Rule : rule system
(
  {Lookup.majorType == system}
)
: label
->
: label.System = {rule = "rule system"}

```

Figure 6. JAPE rule to extract objects.

as an object type, e.g. system, library, etc. We made an extension of Gazetteer lists with new objects related to the context. Then, to extract this concept, we developed a JAPE rule as shown in Figure 6.

4.1.2 Extracting use case concept

In requirement specifications, the action denoted as verbs are considered as use cases. The combination of a verb and a noun (e.g. selects a vehicle) constitutes a use case. Similarly, the adverb (e.g. selection) or a combination of an adverb and a noun (e.g. selection of a vehicle) identifies a use case. We extend Gazetteer lists with new lists of verbs and nouns and we developed JAPE rules to identify verbs, nouns, adverbs, and JAPE rule for the use case extraction.

4.1.3 Extracting include dependency

We use “include dependency” whenever one use case needs the behavior of another. In requirement specifications, we have noticed that generally before an inclusion relation, verbs or phrases whose express the obligation are used (e.g. should and must). We extend Gazetteer lists with new lists of modal verbs and developed a JAPE rule to extract “include dependency”.

4.1.4 Extracting extend dependency

An “extend dependency” is a relationship where an extending use case completes the behavior of a basic use case. Verbs and phrases that express possibility are usually used to define extension relationship (e.g. may and might). We extend Gazetteer lists with new lists of those verbs.

4.1.5 Extracting actor concept

An actor specifies a role played by an external entity that interacts with the system (e.g., by exchanging messages). He can be a human user of the designed system, some other systems or hardware interacting with the system. Actors are generally nouns, e.g. customer, supplier, student, etc. To identify an actor, we identify nouns in the text and we verify in addition if the noun is followed by a use case, since each actor should be associated to one or more use cases.

4.2 Text pre-processing

After transformation rules identification, the text related to user requirements should be preprocessed to detect and tag its components.

The input of this stage is a text related to requirements specification. The text is parsed to have paragraphs, sentences, and tokens. Then, each word (token) is given a grammatical category. This step is performed by applying some GATE API functionalities on user requirements to obtain texts with grammatical categories (i.e. verbs, nouns, ... etc).

The text preprocessing consists of three steps:

- *Sentence splitting*: It consists in breaking down the text into a set of sentences. In this step we use Gate component: *Sentence splitter*.
- *Tokenization*: To split each sentence into tokens, that are words and punctuations. In this step we use Gate component: *Tokenizer*.
- *POS tagging*: Each token, is given the appropriate word category like noun, verb, adjective, etc. In this step we use Gate *Pos Tagger* component. However, during his processing it looks up the Gazetteer lists and JAPE rules to generate the result.

4.3 Execution of transformation rules

This phase involve the extraction of UML use case concepts. Using the results generated by pre-processing of user requirements stage, we apply the transformation rules and extended JAPE rules and Gazetteer lists. This phase is considered as a semantic annotation, allowing the detection and tagging of the concepts of use case diagram and the relationships between these concepts. The result obtained from this phase is an XML file, that the user or the designer can save, print or use it to visualize the result as a use case diagram in the interface of a given CASE (Computer-Aided Software Engineering) tool, such as ArgoUML.

4.4 Graphical representation

In order to transfer XML file into use case diagrams, we use (XSLT) eXtensible Style sheet Language Transformation, and Simple Api for Xml (SAX) to transform the XML file into a Scalable Vector Graphics (SVG) file.

Then, we visualize the use case diagram from the SVG file. We developed a tool in Java that perform the above step that takes user requirements as input and display use case diagrams correspond to the input requirements.

5 EVALUATION

To validate our proposed approach, we have implemented a tool named Use Case Generator (*UCgen*).

We create our own corpus that are gathered from several documents in several areas such as: commercial, educational, ... etc. Following that, we calculated the following effectiveness metrics: recall, precision and overgeneration (Chinchor 1993). The recall, is the ratio of the number of correct generated concepts to the total number of relevant concepts. It is calculated as follows:

$$Recall = \frac{N_{correct}}{N_{correct} + N_{missing}}$$

Precision is the ratio of the number of correct generated concepts to the total number of irrelevant and relevant generated concepts. It is usually expressed as follows:

$$Precision = \frac{N_{correct}}{N_{correct} + N_{incorrect}}$$

Overgeneration metric measures the percentage of the actual generated concepts that were spurious. It is calculated as follows:

$$Overgeneration = \frac{N_{overgenerated}}{N_{correct} + N_{missing}}$$

We have compared our tool with CM-Builder, since it is the closest system to ours. Values are given in Table 1.

We can note that recall and precision of CM-Builder are well below than ABCD tool. Moreover, regarding overgeneration, UCgen tool is more efficient than CM-Builder, since it makes fewer errors in concepts generation.

We have also compared the functionalities of UCgen tool with other available tools that can perform automated analysis of NL requirement specifications. The results of this comparison are given in Table 2.

Table 2 shows that unlike many tools, UCgen tool is able to identify information such as, associations, dependency, ... from natural language requirements. Also, UCgen tool does not need user involvement to detect any concept which makes it a fully automated tool. These results are very encouraging and support very well the approach proposed in this paper.

Table 1. Performance comparison.

	CM-Builder	UCgen
Recall	73%	88%
Precision	66%	90,3%
Overgeneration	62%	29%

Table 2. Concepts generation of UCgen and other tools.

Tool	CM-Builder	NL-OOPS	R-Tool	UCgen
Generalization	No	No	No	Yes
Dependency	No	No	No	Yes
Associations	No	No	No	Yes
Actor	Yes	Yes	Yes	Yes
Use case	Yes	Yes	Yes	Yes
System	No	No	No	Yes

6 CONCLUSION

In this paper, we presented an approach that transforms user requirements into UML use case diagram. The approach consists of four steps: the first step is the definition of the requirements and use case metamodels. The second step is the creation of transformation rules, allowing the transformation of the requirements metamodel into the use case metamodel. The third step is a text processing to tag the text components. Then, the last step is transformation rules execution on the processed text, the result of this step is an XML file, used to generate, automatically, a use case diagram. We developed a tool that support our approach. In future work, we will evaluate the proposed approach and apply it to several case studies. Moreover, we will try to extend the approach to cover other UML diagrams such as: class diagram (Ben Abdesslem Karaa, et al., 2015), sequence diagrams, etc.

REFERENCES

- Afreen, H., Bajwa, I. S., & Bordbar, B. (2011, December). SBVR2UML: A challenging transformation. In *Frontiers of Information Technology (FIT)*, 2011 (pp. 33–38). IEEE.
- Ben Abdesslem Karaa, W., Ben Azzouz, Z., Singh, A., Dey, N., S Ashour, A., & Ben Ghazala, H. (2015). Automatic builder of class diagram (ABCD): an application of UML generation from functional requirements. *Software: Practice and Experience*.
- Chinchor, N. (1993). The statistical significance of the MUC-5 results. In *Proceedings of the 5th Conference on Message Understanding, MUC 1993*, Baltimore, Maryland, USA, August 25–27, 1993, pp. 79–83.
- Cunningham, H., Maynard, D., Bontcheva, K., & Tablan, V. (2002, July). A framework and graphical development environment for robust NLP tools and applications. In *ACL* (pp. 168–175).
- Dube, M. R., & K Dixit, S. (2012). Modeling theories and model transformation scenario for complex system development. *International Journal of Computer Applications*, 38(7), 11–18.
- Harmain, H. M., & Gaizauskas, R. (2003). CM-Builder: A natural language-based CASE tool for object-oriented analysis. *Automated Software Engineering*, 10(2), 157–181.

- Mich, L., & Garigliano, R. (2002). NL-OOPS: A requirements analysis tool based on natural language processing. In Proceedings of Third International Conference on Data Mining Methods and Databases for Engineering, Bologna, Italy.
- OMG (2016). www.omg.org/ / accessed, April 2016.
- Segura, S., Benavides, D., Ruiz-Cortés, A., & Escalona, M. J. (2007). From requirements to web system design. an automated approach using graph transformations. *Actas de Talleres de Ingeniería del Software y Bases de Datos*, 1(6), 61.
- UML (2016). www.omg.org/spec/UML/2.5/. / accessed, April 2016.
- Vinay, S., Shridhar, A., & Prashanth, D. (2009). An Approach towards Automation of Requirements Analysis. Proceedings of the International Multi-Conference of Engineers and Computer Scientists. IMECS 2009. Hong Kong. Vol I. pp. 1–6.

Modeling guidelines of FreeRTOS in Event-B

Eman Alkhamash

Department of Computer Science, University of Taif, Taif, Saudi Arabia

Michael Butler & Corina Cristea

Electronics and Computer Science, University of Southampton, Southampton, UK

ABSTRACT: Formal methods have emerged as an approach to ensuring quality and correctness of highly critical systems. Event-B is a formal method for modelling and reasoning about systems. This paper presents a set of guidelines for modelling OS kernels for embedded real-time systems in Event-B. The presented modelling guidelines are intended to assist specifiers of real-time kernels with a set of modelling steps for the construction of formal models of real-time kernels. Each of these guidelines gives directions on how to model a certain aspect of RTOS kernels. The guidelines focus on the basic functionality of RTOS and represent the primary requirements of RTOS for an Event-B model. Design details that are RTOS-specific are left out from the guidelines. Design details of a specific-RTOS can be specified through a refinement of the abstract model driven by the guidelines. The identified modelling guidelines can be understood as a modelling pattern for the following RTOS features: task management, scheduling and context switch, and interrupts.

1 INTRODUCTION

Engineering cookbooks and guidelines for modeling, refinement, and proofs are important in the context of formal methods. They can be used to systematise the modeling and refinement process and aiding the proofs and therefore reduce the time and cost of the formal development of systems.

Both, new specifiers and professional ones can gain benefit of adopting such guidelines and modeling patterns on their specifications. The guidelines tell how most effectively to model a system and provide good examples and lessons that aid the formal development of several systems. The guidelines also save specifier's time and make formal method approach more acceptable in industry. Moreover, guidelines can shed light into the requirement of systems and draw attention to some important properties of systems that might get missing in the absence of the guidelines. The guidelines may also give insight to the specifier on which requirement should be model first and how to organize the refinement steps as this is usually considered a source of difficulty in the process of modeling and refinement.

This paper provides set of modeling guidelines of RTOS including task management, scheduler states, scheduling and context switching, and interrupts and interrupt service routine. Although, the guidelines provided in this paper are drawing upon our experience with modeling FreeRTOS; it is believed that the presented guidelines can be

applied for constructing formal models of different real-time operating system (RTOS) kernels. This is because RTOS kernels share similar components and features (Li & Yao 2003).

Table 1 describes some of the basic features and concepts of four RTOSes: FreeRTOS (Barry 2010), UCOS (Company 2016), eCos (eCos Company 2016), and VxWorks (River 2016).

This paper is structured as follows. Section 2 gives an overview about Event-B. Section 3 to section 6 identifies respectively modeling concepts for process components, scheduler states, scheduler and context switching, and scheduling and interrupts. Section 7 discusses some related work and finally Section 8 outlines the conclusions.

2 EVENT-B

Event-B (Abrial 2010) is a formal method that uses set theory and first order logic to provide a formal notation for the creation of models of discrete systems and the undertaking of several refinement steps. In the refinement steps, the models represent different abstraction levels of system design. The consistency between the refined models in Event-B is verified through a set of mathematical proof obligations expressing that it is a correct refinement of its abstraction. The complexity in the design process is managed by abstraction and refinement. Refinement allows

Table 1. Summary of the basic features of FreeRTOS, UCOS, eCos, and VxWorks.

Component	FreeRTOS	UCOS	eCos	VxWorks
Process TCB	yes	yes	yes	yes
Null process	yes	yes	yes	yes
Process creation and termination	yes	yes	yes	yes
Process priority	yes	yes	yes	yes
Clock tick	yes	yes	yes	yes
Process states	ready, running, blocked, suspended	dormant, ready, running waiting, ISR	running, sleeping, countsleep, suspended, creating, exited	ready, suspended, pending (blocked) and delayed
Scheduler state	start the scheduler, lock the scheduler, and unlock the scheduler the scheduler, and unock the scheduler	lock the scheduler, and unlock the scheduler	Start the scheduler, lock the scheduler, and unlock the scheduler	Lock the scheduler, and unlock the scheduler unlock the scheduler
Context switch	yes	yes	yes	yes
Interrupts handling	yes	yes	yes	yes
Priority manipulation	changes a process priority and returns the priority of a process	changes a process priority	changes a process priority	changes a process priority and returns the priority of a process a process priority and returns the priority of a process
Delay operation	yes	yes	yes	yes
Scheduling	priority- based round robin scheduling	priority- based scheduling and round robin scheduling	priority based round robin and bitmap schedulers	priority- based scheduling and round robin scheduling
Queue creation and termination	yes	yes	yes	yes
Process synchronization	queues, semaphores, mutexes	semaphores, message mailbox, message queues, tasks and Interrupt service routines (ISR)	message box, semaphore, queue	message queues, pipes, semaphores
Waiting list for tasks waiting to retrieve/post message to a queue	yes	yes	yes	yes
Memory management	memory al- location and deallocation	memory al- location and deallocation	memory allocation	memory al- location and deallocation

to postpone the introduction of some system features to later refinement steps. Rodin (Abrial, Butler, Hallerstede, & Voisin2006, Abrial, Butler, Hallerstede, Hoang, Mehta, & Voisin 2010) is a tool used in the development of Event-B models. Rodin improves the quality of models through many features such as the identification of errors, the identification of required invariants and proofs of consistency.

Event-B notation contains two constructs: a context and a machine. The context is the static part in which we can define the data of the model: sets, constants, axioms that are used to specify assumptions about sets and constants, and theorems that are used to describe properties derivable from the axioms. The dynamic and functional behaviour of a model is represented in the machine part, which includes variables to describe

any *t* where *guard* then *action* end

Figure 1. The outline of an event.

the states of the system, invariants to constrain variables, theorems to describe properties that follow from the invariants, and events to trigger the behaviour of the machine. The outline of an event used in this paper is shown in Figure 1. The event takes parameter *t*, that satisfies the guard and then executes the body. The guard is predicate on the machine variables and event parameters and an action updates machine variables atomically.

3 TASK MANAGEMENT

3.1 Process

A process is an independent thread of execution (Barry 2010, Li & Yao 2003). RTOS can execute multiple processes concurrently (Barry 2010, Li & Yao 2003). The processes appear to execute concurrently, however, kernel interleaves the execution sequentially based on specific scheduling algorithm (Barry 2010, Li & Yao 2003). The term process is also called thread or task in some RTOSes.

To define a process, we define a new type in the context *PROCESS* “carrier set” where each process within kernel is an element of this set.

3.2 Process table

Process table is a data structure consisting of a collection of elements to store the information of a process (Li & Yao 2003, Labrosse 1998). Each process has its own control block that contains the process information such as process id, process priority, ...etc. A process table could consist of more than 10 elements. To deal with process table, we introduce the following concepts in the Event-B model:

Process set: A set of the possible processes available in the system.

Process attributes functions: These are (constant) functions that map processes to the process elements.

Process creation event: An event used to create a process.

Process attributes can be modeled by introducing a constant function for each attribute. For instance, let us use *P* to identify the set of the created processes (i.e $P \subseteq PROCESS$), P_1, P_2, \dots, P_N are sets that define process elements such as name, priority, ...etc., the constant functions pe_1, pe_2, \dots, pe_n can be defined as follows:

$$\begin{aligned} pe_1 &\in P \rightarrow P_1 \\ pe_2 &\in P \rightarrow P_2 \dots \\ pe_n &\in P \rightarrow P_N \end{aligned}$$

The functions correspond to the process attributes do not need to be introduced in one machine, some of the elements can be postponed until a later refinement level depending on what features the specifier wants to model first and what features the specifier wants to postpone.

When a new process is created, the kernel instantiates the process block of the created process. To model this, we introduce the process creation event as follows:

$$\begin{aligned} PC &= \text{anyp where } p \in P \\ &\text{then } Pe_1(P) := v_1 \\ &Pe_2(P) := v_2 \dots \text{end} \end{aligned}$$

The process creation event needs to be extended further in later refinement steps when a new element of a process is introduced.

The definition of process table is similar to the approach to records in Event-B (Evans & Butler 2006). In fact, the carrier set *P* can be thought of as a record type. The attributes pe_1, pe_2, \dots, pe_n are defined using a projection function (function from *P* to some type P_1, P_2, \dots, P_N). PC event are used to create new processes. It is possible to extend the record type *P* by adding more attributes in another refinement step.

3.3 Process priority

In real-time kernels, each process has a priority as defined in the process block. Most RTOS implies a combination of base priority and active priority. The base priority is the original priority specified when the task is constructed. The process priority is the priority that can possibly be modified. One possible usage of base and active priority is in priority inheritance protocol. Priority inheritance protocol takes place when a lower priority process blocks some higher priority processes, this problem called priority inversion. The priority inheritance protocol resolves this problem by raising the priority of the process caused the problem by the highest priority process blocked by it to release the blocked processes. When the process caused the problem releases the blocked processes, its priority then returns back to its base priority. For this, it is important to have a base priority that holds the original priority of the process and active priority that holds the priority that can change. To deal with this, we need to define two elements in the

process block, one for active priority and one for base priority.

$$\begin{aligned} \text{ActivePri} &\in P \rightarrow \text{PRIORITY} \\ \text{BasePri} &\in P \rightarrow \text{PRIORITY} \end{aligned}$$

We need also to introduce the following Event-B events that capture the most common operations related to process priority which are priority set PriSet and priority get PriGet :

$$\begin{aligned} &\text{PriSet} = \text{any } p \text{ } np \\ \text{where } p &\in P \\ &np \in \text{PRIORITY} \\ &\text{ActivePri}(p) \neq np \\ \text{then } &\text{ActivePri}(p) := np \text{ } \text{end} \\ \text{PriGet} &= \text{any } p \text{ } np! \text{ where } p \in . \\ &np! = \text{ActivePri}(p) \text{ } \text{end} \end{aligned}$$

We use the !convention to represent result parameters as shown in the $np!$ parameter.

3.4 Process states

At any time, the process can be in one state (Barry 2010, Labrosse 1998). There are different states of a process, for instance, in FreeRTOS a process can be in one of the following states: ready, suspended, blocked, and running.

State diagram is always useful to show the transition of process states. Figure 1 shows a possible transitions among the states RUNNING , READY , SUSPENDED and BLOCKED during a process life.

To deal with process states, we identify the following concepts in the Event-B modelling:

Process states variables: These variables are defined for each possible state.

One way of defining process variables states in an Event-B model is to have a set for each state, each set is disjoint from other sets indicating that a process must be in one state at any given time. Each created process must belong to one of these

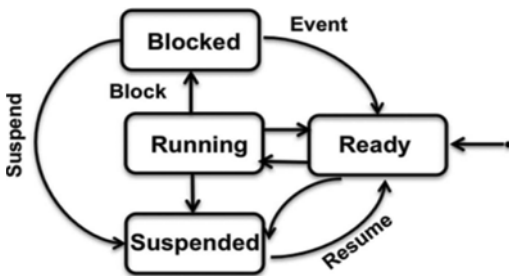


Figure 2. Process states.

sets. We use the partition operation to indicate that the collection of the sets is disjoint.

$$\text{partition}(P, \text{READY}, \text{RUNNING}, \text{SUSPENDED}, \text{BLOCKED})$$

To check the state of a particular process, we check which set the process belongs to.

Running set RUNNING in our diagram corresponds to the runnable process and demands a special treatment. It should be defined as a singleton set if we are dealing with one core processor. Thus, we need to add the following invariant:

$$\text{card}(\text{RUNNING}) \leq 1$$

Process states event(s): These events are used to update the state of the process to a new one. Process states events model the transitions between states. According to our diagram, we have four process states events: RunToReady , RunToBlock , RunToSuspend , ReadyToRun , ReadyToSuspend , BlockedToReady , BlockedToSuspend , and SuspendToReady .

The definition of RunToReady would be of the form: $\text{RunToReady} = \text{any } p$

$$\begin{aligned} &\text{where } p \in \text{RUNNING} \\ \text{then } &\text{RUNNING} := \emptyset \\ \text{READY} &:= \text{READY} \cup \{p\} \text{ } \text{end} \end{aligned}$$

In Run To Ready event, the process p is moved from RUNNING set to READY set.

In later refinement levels, process state variables “sets” are refined to the appropriate data structure. For instance, ready state is usually implemented as a linked list for each priority. This level of detail can be left while introducing data refinement levels. In data refinement levels, sets are data refined to the appropriate data structure such arrays or linked list depending on the data structure implementation adopted for the specified RTOS. For instance, the data structure for process states in FreeRTOS are circular-doubly linked list (Barry 2010).

3.5 Null process

Null or idle process is a special process that is run when there is no process available to run (Barry 2010, Labrosse 1998). Null process is permanently ready to run and is assigned to the lowest priority.

It is an important process as the kernel needs always to execute a process or interrupt handler, and there might be a situation where there is no ready process available to run, in this case, the kernel switch control to the null process.

The definition of the null process is:

$$\text{Null} \in \text{READY}$$

3.6 Timing behaviour

The scheduler interrupts at regular frequency to schedule tasks. RTOS measures time using a tick count variable. A timer increments the tick count with a strict time. Each time the tick is incremented the scheduler checks to see if a task needs to be woken. If this is the case, the scheduler executes the newly woken task.

One important operation used by many RTOSs is *Delay* operation. It is used to suspend a process until a fixed point in the future. *Delay* operation takes a process and a time as parameters and returns the process delayed until the specified delay time.

To deal with timing behaviour, we identify the following Event-B modeling concepts:

Tick variable: A counter that measures the system time ($Tick \in N$).

Tick length constant: A constant that defines the length of the tick depending on the hardware timer's design.

Sleep time function: Partial function that maps processes to sleep time. The partial function indicates that only some processes in the system have a delay time and not all processes ($Sleep \in P \rightarrow N$).

Time To Wake function: Partial function describes the wake-up time of a delayed task ($TimeToWake \in BLOCKED \rightarrow N$) where *BLOCKED* identifies the set of delayed processes. Since all processes that have wake-up time are also have sleep time, we add the following invariant:

$$dom(TimeToWake) \subseteq dom(Sleep)$$

Increment Tick: An event used to increment a tick counter with a strict accuracy.

IncrementTick = where $Tick \geq 0$
 $\wedge ((TimeToWake \neq \emptyset \wedge \min(ran(TimeToWake)) > Tick) \vee TimeToWake = \emptyset)$
then $Tick := Tick + TickLength$ **end**

The guard $TimeToWake \neq \emptyset \wedge \min(ran(TimeToWake)) > Tick$ ensures that all delayed processes have un-expired delay time. This is because, each time a tick count is incremented it must check if a delayed task needs to be woken.

Delay: An event used to delay a process for a fixed time.

Delay = any p
where $p \in RUNNING$
 $Sleep(p) > 0$
then $RUNNING := \emptyset$
 $BLOCKED := BLOCKED \cup \{p\}$
 $TimeToWake(p) := Sleep(p) + Tick$ **end**

4 SCHEDULER STATES

The scheduler can exist in one of different states (Barry 2010, Labrosse 1998). For instance, the possible scheduler states in FreeRTOS are: not-started, running, and suspended.

To deal with scheduler states, we have:

Scheduler states constants: Constants represent scheduler states.

Scheduler status variable: A variable whose value determines the state of the scheduler.

Scheduler events: These events are used to update the value of the scheduler variable to a new one.

Assuming that we have a scheduler with three states s_1, s_2 and s_3 . We use *ss* for scheduler variable, s_1, s_2 and s_3 are constants corresponding to the scheduler states that are defined in the context level and SE for a scheduler event. To define scheduler states, partition operation can be used as follows:

Partition($ss, \{s_1\}, \{s_2\}, \{s_3\}$)

A scheduler event would be of the form:

SE = when $ss = s_1$ **then** $ss = s_2$ **end**

The conditions of which the transition occurred are considered as guards for the scheduler states events.

There might be more than one event that changes the value of the scheduler variable and there all would take the aforementioned form.

Many operations in RTOS are occurred when the scheduler is running or suspended, for instance, *IncrementTick* event given in section 3.6 must occurs in FreeRTOS when the scheduler runs; therefore *IncrementTick* event is extended by adding the guard $ss = s_1$ where s_1 denotes the scheduler running state.

5 SCHEDULING AND CONTEXT SWITCHING

5.1 Context switch

A context switch occurs when the scheduler switches from one process to another. It is the process of replacing the process being executed by another process that is ready to run (Li & Yao 2003, Labrosse 1998). Context switch takes place when the kernel decides to switch control to other process, so, it exchanges the registers contents by saving the information of the current executing process to resume its execution later and loading the information of the new process to the processor registers.

There are many situations in which context switch is performed. The most common context

switch performed is tick interrupt. At every tick, the scheduler checks if a new process should run. If such process is found, the scheduler will save the current process information to resume it later and execute the other one (Barry 2010).

The kernel schedule processes based on a specific algorithm. There are different scheduling algorithms; a preemptive priority scheduling mechanism is one of the most common scheduling algorithms. In this algorithm, each process has a priority, and the higher readied priority process runs first. The preemptive priority scheduling algorithm can augment round robin algorithm by giving processes of the same priority an equal share of processor time.

To deal with context switch, we introduce the following modeling concepts:

Context function: A total function that maps each task to its context (attribute in the process table)

$$(ProcessContext \in P \rightarrow CONTEXT).$$

Current context: The physical context that captures the current process that holds the processor.

$$PhysicalContext \subseteq CONTEXT$$

$$card(physicalContext) \leq 1$$

Context switch event(s): Event(s) replace the current process by another one.

These guidelines are dealing with preemptive priority scheduling algorithm. At every tick, a context switch is performed when there is a process with a priority higher than or with equal priority of the current one being executing (equal in this context allows time slicing between processes with same priorities). The guards or the conditions that constrain the context switch CS event rely on checking the priority of the readied processes, thus, the context switch event can be of the form:

$$\begin{aligned} & \text{ContextSwitch} = \text{any } p \text{ } cp \text{ } ph \text{ where } p \in READY \\ & cp \in RUNNING \\ & \quad activePri(p) \geq activePri(cp) \\ & \quad ph \in physicalContext \\ & \text{then } RUNNING := \{p\} \\ & \quad physicalContext := \{ProcessContext(p)\} \\ & \quad ProcessContext(cp) := ph \text{ end} \end{aligned}$$

The actual context switch is done through saving the state of the current context in the process context and set the current context for the new running process.

Context switch flag: A Boolean flag that set to true to indicate the need of performing a context switch. We use *csFlg* for context switch flag.

There are different cases where a context switch may perform. For instance, context switch might

be performed when a certain process transmits to ready state or when a new process is created, because it is possible for these situations to introduce a process with a priority higher than the priority of the current executing one in which case the context switch should perform. Therefore, for every event that may require a context switch, we need to add an additional action that set the context switch flag to true. And we add an additional guard to ensure that the context switch flag is true in the context switch CS event, and finally the context switch flag is assigned back to false in the CS event. Thus, we extend the CS event to be:

$$\begin{aligned} & \text{ContextSwitch} = \text{any } p \text{ } cp \text{ } ph \\ & \quad \text{where } p \in READY \\ & \quad cp \in RUNNING \\ & \quad activePri(p) \geq activePri(cp) \\ & \quad ph \in physicalContext \\ & csFlg = TRUE \\ & \quad \text{then } RUNNING := \{p\} \\ & \quad physicalContext := \{ProcessContext(p)\} \\ & \quad ProcessContext(cp) := ph \\ & csFlg := FALSE \text{ end} \end{aligned}$$

In order to prevent any event to be enabled between calling the context switch event, we add an additional guard *csFlg = FALSE* to any events that might get enabled and affect the execution of the context switch.

Another important issue raised here is what if an event calls a context switch and sets the *csFlg* to true but not all the guards of the context switch CS event are satisfied. This situation occurs when the event that called a context switch does not introduce a readied process with higher or equal priority than the executing one. This means that context switch event is not enabled and consequently the value of *csFlg* is always true which leads to a deadlock. To prevent this situation, we suggest to have two versions of context switch events, the first one performs the context switch if all the guards of CS event are satisfied, the second one, however, triggers only to set *csFlg* back to false when the guards of the context switch are not satisfied to prevent a possible deadlock from occurring. The revised CS events are:

$$\begin{aligned} & \text{ContextSwitch } 1 = \text{any } p \text{ } cp \text{ } ph \\ & \quad \text{where } p \in READY \\ & \quad cp \in RUNNING \\ & \quad activePri(p) \geq activePri(cp) \\ & \quad ph \in physicalContext \\ & csFlg = TRUE \\ & \quad \text{then } RUNNING := \{p\} \\ & \quad physicalContext := \{ProcessContext(p)\} \\ & \quad ProcessContext(cp) := ph \\ & csFlg := FALSE \text{ end} \end{aligned}$$

ContextSwitch2 = *any* *cp*
where $cp \in RUNNING$
 $\max(activePri[READY]) < activePri(cp)$
 $csFlg = TRUE$
then $csFlg := FALSE$ *end*

The above context switch specification is abstract and only represents the general operations performed in any context switching. Since context switching is a hardware-dependent operation, in order to model context switch in details, the registers of the target processor needs to be modeled along with the other specific context switch details for the targeted architecture.

6 INTERRUPTS AND INTERRUPT SERVICE ROUTINES

Interrupts are hardware mechanisms used to inform the kernel that an event has occurred (Barry 2010, Labrosse 1998). A process can be interrupted by external interrupts raised by peripherals or software interrupts raised by executing a particular instruction. Interrupts are handled by ISRs which are stored in interrupt vector table. When an interrupt occurs, the kernel saves the current context of the process being interrupted and jumps to the ISR to handle that interrupt. After processing the ISR, the kernel returns to the process level and resumes the process was interrupted (Barry 2010, Labrosse 1998).

Interrupts may ready a blocked process for an external device when an event has occurred, so, the kernel might execute different process before completing the preempted one.

To deal with interrupts, we introduce the following modeling concepts:

Interrupt data type: A new data type for interrupts. We will use *INTERRUPT* “carrier set” as interrupts data type.

Interrupt variable: A set of the raised system interrupts $Interrupts \subseteq INTERRUPT$.

Interrupt handler function: Function that maps each interrupt to its ISR. We will use *interrupt handler* (a total function that maps an interrupt to its corresponding interrupt service routine).

$ISR \in Interrupts \rightarrow INTERRUPT\ HANDLER$

Current interrupt: A variable that stores the current executing interrupt.

$currentISR \subseteq ran(ISR)$
 $card(currentISR) \leq 1$

$card(currentISR) \leq 1$ is added in case of single kernel.

Handle interrupt event: An event used to handle interrupts.

Handle_Interrupt = *any* *int*
where $int \in Interrupts$
then $currentISR := \{ISR(int)\}$ *end*

Complete interrupt event: an event used to discard completed interrupts

Complete_Interrupt = *any* *int*
where $int \in Interrupts$
 $currentISR = \{ISR(int)\}$
then $currentISR := \emptyset$
 $Interrupts := Interrupts \setminus \{int\}$
 $ISR := \{int\} \leftarrow ISR$ *end*

For later refinement steps, more features that can be introduced to be used from within an ISR and we need to make a distinction between the features used by process level and the features used by interrupt level. For instance, if we were to introduce an event for sending an item to a queue that is used by ISRs and processes, we need to introduce two events, Queue- SendFromISR is an event used by ISRs and Queue- SendProcess is an event used by processes.

Interrupts have priority; an interrupt with lower priority can be interrupted by other interrupt of a higher priority. To deal with this, we identify the following Event-B modeling concepts:

Interrupt priority variable: A function that maps each interrupt to its priority.

$InterruptPriority \in Interrupts \rightarrow INTERRUPT_PRIORITY$

Where *INTERRUPT PRIORITY* is a constant set defined in the context level.

The event *Handle Interrupt* and *Complete Interrupt* can be extended as follows:

Handle_Interrupt = *any* *int* *c* *cint*
where $int \in Interrupts$
 $(c \in currentISR \wedge cint = ISR^{-1}(c))$
 $\wedge InterruptPriority(int) \geq InterruptPriority(cint) \vee$
 $currentISR = \emptyset$
then $currentISR := \{ISR(int)\}$ *end*

Complete_Interrupt = *any* *int*
where $int \in Interrupts$
 $currentISR = \{ISR(int)\}$
then $currentISR := \emptyset$
 $Interrupts := Interrupts \setminus \{int\}$
 $ISR := \{int\} \leftarrow ISR$
 $InterruptPriority := \{int\} \leftarrow InterruptPriority$ *end*

The processor always gives priority to execute interrupts over tasks, the ISR must complete its execution without being interrupted by tasks. When the ISR is completed, the kernel dispatches the correct task. To deal with this, we identify the following modeling concepts:

Interrupt context: A function that maps each interrupt to its context.

$$\text{InterruptContext} \in \text{Interrupts} \rightarrow \text{CONTEXT}$$

In order to separate a process context from an interrupt context; we add the following invariant:

$$\text{ran}(\text{InterruptContext}) \cap \text{ran}(\text{ProcessContext}) := \emptyset$$

The *Handle Interrupt* and *Complete Interrupt* are extended as follows:

Handle Interrupt = *any* *int* *c* *cint* *ph*
where *int* \in *Interrupts*
 $c \in \text{currentISR} \wedge \text{cint} = \text{ISR}^{-1}(c)$
 $\wedge \text{InterruptPriority}(\text{int}) \geq \text{InterruptPriority}(\text{cint}) \vee$
 $\text{currentISR} = \emptyset \wedge$
 $\text{ph} \in \text{physicalContext}$
then $\text{currentISR} := \{\text{ISR}(\text{int})\}$
 $\text{physicalContext} := \{\text{InterruptContext}(\text{int})\}$
 $\text{InterruptContext}(c) := \text{phend}$
Complete Interrupt = *any* *int*
where *int* \in *Interrupts*
 $\text{currentISR} = \{\text{ISR}(\text{int})\}$
then $\text{currentISR} := \emptyset$
 $\text{Interrupts} := \text{Interrupts} \setminus \{\text{int}\}$
 $\text{ISR} := \{\text{int}\} \leftarrow \text{ISR}$
 $\text{InterruptPriority} := \{\text{int}\} \leftarrow \text{InterruptPriority}$
 $\text{InterruptContext} := \{\text{int}\} \leftarrow \text{InterruptContext}$
 $\text{physicalContext} := \emptyset \text{end}$

An extra guard $\text{currentISR} = \emptyset$ is needed in CS events to prevent any context switch while there is an ISR running.

The timer interrupt given in section 3.6 is an example of an interrupt. In every tick, the tick-ISR represented in IncrementTick event wakes up the blocked process that have expired delay time. If the woken process has a priority higher than the current process, the ISR then will return control to the higher priority process.

7 RELATED WORK

This section examines some of the related work regarding the use of formal methods in operating systems.

Craig's work is one of the fundamental sources in this field (Craig 2007a, Craig 2007b). He focuses on the use of formal methods in OS development, and the work is introduced in two books. The books contain formal specifications of simple and separation kernels along with the proofs written by hand. The first book is dedicated to specify the common structures in operating system kernels in Z (Spivey 1992) and Object Z (Smith 2000), with some CCS (Milner 1989) (Calculus of Communicating Systems) process algebra used to describe the hardware operations. It starts with a simple kernel with few features and progresses on to more complex examples with more features. For example, the first specification introduced in the book is called a simple kernel, and involves features such as task creation and destruction, message queues and semaphore tables. However, it does not contain a clock process or memory management modules, whereas other specifications of swapping kernel contain more advanced features including a storage management mechanism, clock, interrupt service routines, etc.

The second book is devoted to the refinement of two kernels, a small kernel and a micro kernel for cryptographic applications. The books contain proofs written by hand and some missing properties resulting due to manual proofs, which have been highlighted by Freitas (Freitas 2009).

Freitas (Freitas 2009, Velykis & Freitas 2010) has used Craig's work to explore the mechanisation of the formal specification of several kernels constructed by Craig using Z/Eves theorem prover. This covers the mechanisation of the basic kernel components such as the process table, queue, and round robin scheduler in Z. The work contains an improvement of Craig's scheduler specification, adapting some parts of Craig's models and enhancing it by adding new properties. New general lemmas and preconditions are also added to aid the mechanisation of kernel scheduler and priority queue. Mistakes have been corrected in constraints and data types for the sake of making the proofs much easier, for instance, the enqueue operation in Craig's model preserves priority ordering, but it does not preserve FIFO ordering within elements with equal priority; this has been corrected by Freitas in (Freitas 2009).

Furthermore, Déharbe et al (Déharbe, Galvao, & Moreira 2009b) specify task management, queues, and semaphores in classical B. The work specifies mutexes and adopts some fairness requirements to the scheduling specification. The formal model built was published in (Déharbe, Galvao, & Moreira 2009a).

There is also an earlier effort by Neumann et al (Neumann, Boyer, Feiertag, Levitt, & Robinson 1980) to formally specify PSOS (Provably Secure

Operating System) using a language called SPECIAL (SPECification and Assertion Language) (Feiertag & Neumann 1979). This language is based on the modelling approach of Hierarchical Development Methodology (HDM). In this approach, the system is decomposed into a hierarchy of abstract machines; a machine is further decomposed into modules, each module is specified using SPECIAL. Abstract implementation of the operations of each module are performed and then is transformed to efficient executable programmes. The work began in 1973 and the final design was presented in 1980 (Neumann, Boyer, Feiertag, Levitt, & Robinson 1980). PSOS was focusing on the kernel design and it was unclear how much of it has been implemented (der Rieden 2009). Yet, there are other works inspired by the RSOS design such as KernelizedSecure Operating System (KSOS) (Perrine, Codd, & Hardy 1984) and the Logical Coprocessing Kernel (LOCK) (Saydjari, Beckman, & Leaman 1987).

The aforementioned examples follow a top-down formal method approach, where the specification is refined stepwise into the final product. On the other hand, there are also some earlier efforts in the area of formal specification and correctness proofs of kernels based on the bottom-up verification approach. The bottom-up approach adopts program verification methods to verify the implementation.

An example of this approach is a work by Walker et al (1980) (Walker, Kemmerer, & Popek 1980) on the formalisation of the UCLA Unix security kernel. The work is developed at the University of California at Los Angeles UCLA for the DEC PDP-11/45 computer. The kernel was implemented in Pascal due to its suitability for low-level system implementation and the clear formal semantics (Hoare & Wirth 1973, Radha 1999). Four levels of specification for the security proof of the kernel were conducted. The specifications were ranging from Pascal code at the bottom to the top-level security properties. After that, the verification based on the first-order predicate calculus was applied that involves the proof of consistency of different levels of abstraction with each other. Yet, the verification was not completed for all components of the kernel.

Finally, there was an effort by Klein et al (Klein, Elphinstone, Heiser, Andronick, Cock, Derrin, Elka-duwe, Engelhardt, Kolanski, Norrish, Sewell, Tuch, & Winwood 2009, Klein, Derrin, & Elphinstone 2009) on the formal verification of the seL4 kernel starting with the abstract specification in higher-order logic, and finishing with its C implementation. The design approach is based on using the functional programming language Haskell (Hudak, Peterson, & Fasel 2000) that provides an

intermediate level that satisfies bottom-up and top-down approaches by providing a programming language for kernels developer and at the same time providing an artefact that can be automatically translated into the theorem prover. A formal model and C implementation are generated from seL4 prototype designed in Haskell. The verification in Isabelle/HOL (Nipkow, Wenzel, & Paulson 2002) shows that the implementation conforms with the abstract specification.

In this paper, the modeling guidelines adopts top-down formal method approach. The guidelines are modeled in Event-B which is a refinement based approach for modeling systems. The guidelines outlined in this paper can be refined later to the appropriate data refinement structure such as linked list.

8 CONCLUSIONS AND FUTURE WORK

In this paper, we developed modeling guidelines in Event-B for the following RTOS features: task management, scheduling and context switch, and interrupt and interrupt service routines. The devised guidelines need to be evaluated through several case studies. As a direction for future research, we are going to evaluate and extend the guidelines by applying them to particular RTOS. We expect to improve these guidelines and cover some more guidelines on queue management and memory management.

REFERENCES

- Abrial, J.-R. (2010). *Modeling in Event-B—System and Software Engineering*. Cambridge University Press.
- Abrial, J.-R., Butler, M., Hallerstede, S. & Voisin, L. (2006). An open extensible tool environment for Event-B. In *ICFEM 2006, LNCS*, pp. 588–605. Springer.
- Abrial, J.-R., Butler, M., Hallerstede, S., Hoang, T., Mehta, F. & Voisin, L. (2010). Rodin: an open toolset for modelling and reasoning in Event-B. *STTT* 12(6), 447–466.
- Barry, R. (2010). The FreeRTOS project. <http://www.freertos.org/>.
- Company, M. (2016). Micrium embedded software. <http://ecos.sourceware.org/>.
- Craig, I. (2007a). *Formal models of operating system kernels*. Springer.
- Craig, I. (2007b). *Formal Refinement for Operating System Kernels*. Secaucus, NJ, USA: Springer-Verlag New York, Inc.
- Deharbe, D., Galvao, S. & Moreira, A. (2009a). <http://code.google.com/p/freertosb/source/browse>.
- Deharbe, D., Galvao, S. & Moreira, A. (2009b). Formalizing FreeRTOS: First steps. In Oliveira M. V. M. and Woodcock, J. (Eds.), *SBMF, Volume 5902 of Lecture Notes in Computer Science*, pp. 101–117. Springer.

- derRieden, T. (2009). Verified Linking for Modular Kernel Verification.
- eCos Company. (2016). eCos. <https://www.micrium.com/rtos/ucosii/overview/>.
- Evans, N. & Butler, M. (2006). A proposal for records in Event-B. In Formal Methods 2006, McMaster, Canada pp. 221–235.
- Feiertag, R. & Neumann, P. (1979). The foundations of a provably secure operating system (PSOS). In IN PROCEEDINGS OF THE NATIONAL COMPUTER CONFERENCE, pp. 329–334. AFIPS Press.
- Feiertag, R. J., Levitt, K. N. & Robinson, L. (1980). A provably secure operating system: The system, its applications, and proofs. In Technical Report CSL-116, SRI International.
- Freitas, L. (2009). Mechanising data-types for kernel design in Z. In SBMF, pp. 186–203.
- Hoare, C. & Wirth, N. (1973, December). An axiomatic definition of the programming language PASCAL. *Acta Informatica* 2(4), 335–355.
- Hudak, P., Peterson, J. & Fasel, J. (2000). A gentle introduction to Haskell, haskell.org.
- Klein, G., Derrin, P. & Elphinstone, K. (2009, August). Experience report: sel4: formally verifying a high-performance microkernel. *SIGPLAN Not.* 44(9), 91–96.
- Klein, G., Elphinstone, K., Heiser, G., Andronick, J., Cock, D., Derrin, P., Elkaduwe, D., Engelhardt, K., Kolanski, R., Norrish, M., Sewell, T., Tuch, H. & Winwood S. (2009). sel4: Formal verification of an OS kernel. In *ACM SYMPOSIUM ON OPERATING SYSTEMS PRINCIPLES*, pp. 207–220. ACM.
- Labrosse, J. (1998). *MicroC/OS-II* (2nd ed.). R & D Books.
- Li, Q. & C. Yao. (2003). *Real-Time Concepts for Embedded Systems*. CMP Books.
- Milner, R. (1989). *Communication and concurrency*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc.
- Neumann, P. R. Boyer, R.
- Nipkow, T., Wenzel, M. & Paulson, L. (2002). *Isabelle/HOL: a proof assistant for higher-order logic*. Berlin, Heidelberg: Springer-Verlag.
- Perrine, T., Codd, J. & Hardy, B. (1984). An overview of the kernelized secure operating system (KSOS). In *In Proceedings of the Seventh DoD/NBS Computer Security Initiative Conference*, pp. 146–160.
- Radha, G. (1999). *Pascal Programming*. New Age International (p) Limited.
- River, W. (2016). *VxWorks*. <http://windriver.com/products/vxworks/>.
- Saydjari, S., Beckman, J. & Leaman, J. (1987). Locking computers securely. In *In 10th National Computer Security Conference*, pp. 129–141.
- Smith, G. (2000). *The Object-Z specification language*. Norwell, MA, USA: Kluwer Academic Publishers.
- Spivey, M. (1992). *Z Notation—a reference manual* (2. ed.). Prentice Hall International Series in Computer Science. Prentice Hall.
- Velykis, A. & Freitas, L. (2010). Formal modelling of separation kernel components. In *Proceedings of the 7th International colloquium conference on Theoretical aspects of computing, ICTAC'10*, Berlin, Heidelberg, pp. 230–244. Springer-Verlag.
- Walker, J., Kemmerer, A. & Popek, J. (1980, February). Specification and verification of the UCLA unix security kernel. *Commun. ACM* 23, 118–131.

Comparative study of various open source digital content management software

Fudhah Ateq Al Selami

Department of Management Information Systems, College of Business-Alkamel Branch, University of Jeddah, Jeddah, Saudi Arabia

ABSTRACT: The web content management systems have grown in importance as more and more organizations communicate and publish their information via the web. The main objective of this article is to provide a brief directory to research and scientific institutions to enable them to make the right decision in choosing the right system for hosting their digital collections. The researcher chose three systems of open source digital repositories; DSpace, Greenstone, and EPrints, to find out the technical and functional capabilities of them. The study used comparative analytical survey method, and found that the DSpace software is the most widely accepted among digital repositories' solutions. Greenstone and EPrints programs are also used on a large scale as they are two low cost options. Also, we concluded that EPrints meets the demand for technical support and training in the use of digital repositories.

Keywords: digital library software, digital repository, DSpace, greenstone and EPrints

1 INTRODUCTION

In the past years, open access emerged as a key development in the world of scientific communication. A number of initiatives emerged in the world that encouraged free availability. On the top of them is "Budapest open access initiative 2002". "By 'open access' to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited" [1].

Digital repositories are counted among the open access strategies specified in Budapest Initiative. These strategies are called Self Archiving or Green road and consist of databases available on the direct line that provide open access to various types of intellectual production. The digital repositories are also considered among digital collection management systems that are designed to organize and display existing digital content in several forms, and usually owned by somebody, and not licensed from a third party. Most of these

systems are provided by well-known systems suppliers in the field of libraries and these systems rely on a set of standards including XML, metadata protocols (OAI-PMH) and other unified standards [2]. Thus, the digital repositories are a mechanism for the management of digital content and include a wide range of various sources of information for a variety of purposes. The open source software for digital content management divided into two categories of software. The first category is designed for the creation and management of libraries and digital repositories such as Greenstone, EPrints, and DSpace. The second category is designed to manage web content such as Joomla and Drupal.

In this study the researcher choose widely used open source software designed to create and manage digital content that are; Greenstone, EPrints and DSpace. Open source software is a technical program developed by specialists in programming and technology to help create free software solutions. The open source software focus on giving freedom to the users to run, copy, distribute, study, making changes and improve the software. Therefore, the open source software has been built on the following four standard philosophy of the concept of freedom [3]. They are: use for any purpose, no restrictions on the programs, the source and the distribution of amended program. Open source free integrated library systems are more efficient in terms of cost and users' satisfaction than their

commercial counterparts [4]. These systems are distinguished with the following features:

- Capability of design or adaptation to suit local needs: Availability of source code means that the user can edit and develop programs to fit the specific needs unlike commercial products. Also, the development priorities are set by the user, not the vendor.
- No restrictions on its use: there is no contractual restriction on how to use the program unlike commercial software. Developers use them under public-use license that ensures right to modification and distribution for users.
- Low cost: There is no cost for the program itself, and the only major cost is for training and local development cost, which vary from one library to another depending on the needs of libraries and services they wish to provide to their visitors.
- Gaining high percentage of benefits when the cost compared to the benefits achieved.

There are multiple systems to build digital repositories and they vary in their capabilities and specifications. Choosing a system for the repository is the most important step to build digital repositories. Table 1 shows the most important systems of open source digital repositories which are the subject of this study [5–7].

The author assures and confirm that the digital repositories characterized by a set of features and characteristics driven from the nature of the functions. These characteristics distinguish digital repositories from other materials and digital resources available on the web and these characteristics include the following important facilities:

- Object model: It consists of digital content (data) and what is beyond the data (metadata) and Digital Object Identifier (DOI), which identifies the entities in the digital environment. DOI facilitates management of access to entities and is used for archiving and preservation of digital documents for long periods.
- Supporting collections and relations: by this characteristic we mean collecting descriptive

metadata and defining sub collections and templates that describe the digital entities. We also mean by this feature displaying the collection and defining relations between the entities of the same or different types.

- Storing digital content and metadata: storage capacity is mentioned along with the preservation issues. It is important for the digital repository system to guarantee standard as long as the user selects metadata collections and multiple formats for digital content.
- Interoperability: this characteristic is represented in standards that support digital repository systems for ensuring interoperability with other systems as well as export digital entities in the format of important open standards.
- User interfaces: user interfaces are provided for end-user's access to the digital library and its digital collections and entities.
- Access control: this characteristic supports users, collections, methods of verification of identity, license to use and the level of adherence to access and update (digital library collection, digital entity and digital content).
- Search and Browsing: there are mechanisms used for indexing and searching in metadata. It is important for the digital repository system that it should support not only restricted metadata collection but also the specific fields of metadata.
- Entity management: By entity management we mean methods and user interfaces provided by digital repositories system to handle data input, update, deletion, metadata and digital content.
- Multiple language support: it is necessary to support multiple languages in user interface, metadata, and digital content. Also, encoding feature is of great importance to make digital libraries systems a complete multiple languages system.
- Level of customization: by this we mean customizing the digital repository system on the basis of collections, digital entities formats, and services provided. In addition to quality and methods provided by the Application Programming Interfaces (APIs) for digital repositories systems.

Table 1. Important open source digital repositories.

System	Developer or university	License for use	Website
Greenstone	Project of New Zealand digital library in Waikato University, and distributed in association with UNESCO and Human Info NGO.	GNU General Public License	http://www.Greenstone.org/
DSpace	Institute of Mashyshus for Science and Technology and HP company	BSD Open source license	http://www.dSpace.org/
EPrints	Southampton University	GNU General Public License	http://www.EPrints.org/

2 LITERATURE REVIEW

Open source software modes provide the source code free of cost for the user who is working on to adjust the software to suit the needs of his foundation. The digital library software does not ask for cost like commercial software, and enables libraries to control its digital collections and environment work. This study has provided evaluation of the three systems and mentioned its characteristics and usability for digital library management. The researcher applied methodology of documents study relating to software and manuals of technical programs [8].

A large number of digital repositories and digital libraries systems spread as open source software. Their diversity became a disturbing factor for organizations that were planning to build a digital repository to host their collections and simplify the decision-making process for them. The study intended to compare five systems which are: DSpace, Fedora, Greenstone, EPrints and Invenio. The researchers suggested some appropriate systems for different situations of the digital content. These systems were represented in the study of different situations of digital and functional content to guide the organization to choose the appropriate software for the repository and its digital content [9].

A study aimed to identify the different kinds of open source software available for libraries especially open source systems for library management, by examining 13 automated library management system. The study developed some standards that can be used to choose the appropriate system to meet the needs of Arab libraries. The study used survey method to identify the systems currently available on the internet, and analytical method to find out a list of criteria to evaluate these systems. The study found a list of 115 essential criteria that can be used in order to choose an appropriate open source system and can be relied upon in the Arab library management [10].

A study on digital content management and its systems tried to collect digital projects in the world especially in Arab. The researcher evaluated DSpace and Greenstone to see their potential and identify the most appropriate system for managing digital content for Arab information centers. The study used experimental, comparative, evaluative and survey methods and found that DSpace is the world's most widely used system. Also, the study found that there is poor participation from Egypt in the use of digital content management systems. As well as the study found that English is the main language for providing digital content in the world [11].

A study aimed to explain the importance of digital repositories and their role in the conservation

and organizing digital content, as well as to identify the most important international standards for assessing systems designed to build digital repositories for the purposes of preference between them in order to give an opportunity for information institutions to choose the most appropriate of them [12].

It seems good to indicate here that the attention of previous studies was focused on providing criteria for evaluating open source digital repositories software. Also, these previous studies tried to evaluate these software according to the criteria set for each study. This comes in consistent with the purpose of the current study as it aimed to evaluate digital repositories' software (Greenstone, EPrints, and DSpace) according a checklist of basic characteristics of digital repository software that helps assess digital repositories to facilitate for the organization to decide about choosing the appropriate software.

3 THE RESEARCH PROBLEM

The digital repositories in general and institutional repositories in particular are the focus of the study and the research. The researcher has tried to shed light on the concept and importance of open source software for institutional repositories, with a focus on software designed to build these repositories (Greenstone, EPrints, and DSpace). This study aimed to identify the technical and functional capabilities of these software. Although there are many programs available in the market and we can choose from them, the comparative and evaluative studies will help the institutions to choose the program best suited to their digital content.

4 THE RESEARCH METHODOLOGY

The study employed comparative analytical survey method and carried out a survey of free access repositories' directory "Open DOAR" [13]. In this study, the authors pointed out that the DSpace is the most widely used system by 43.5%, followed by EPrints by 13.7%. Also, the study contains 52 repositories available on Greenstone. Our study carried out a survey of software websites available on the internet and read all the literature of the subject in databases and on websites. This is with a view to draw out a checklist of basic characteristics of digital repository software that helps to their evaluation. This list is not comprehensive, but includes most of the necessary characteristics for the management of digital repositories and to compare software related to them.

5 THE MAJOR FINDINGS

Digital repositories systems rely on depositing digital sources through registering function which defines ways to send scientific sources related to researchers or concerned associates. As well as the digital repositories systems rely on building a repository in line with international standards such as the standard of “Open Archives Initiative”. This is to ensure search in the repository content through engines and search tools, and inform the beneficiaries of new materials available in their areas of interest such as RSS service, or preparing a list of new additions and updates. Digital repositories depend on archives for long-term preservation of the works deposited by workers affiliated to the foundation. This happens by applying the best techniques, standards, and well known programs of preservation. Table 2 shows the most prominent technical and functional capabilities of the following digital repositories software systems: DSpace, EPrints and Greenstone.

Comparison of digital repositories systems are made in accordance with the basic characteristics of digital repository as follows:

- Object model: The basic entity of Greenstone is the (document). The structure and content of the document encoding language are formed in XML. The documents are linked with a single or multiple sources, which represents the digital content of the entity. Each document contains a specific identifier, but there is no support for permanent identifiers of the sources. The basic entity of DSpace system is the component (material) containing digital content and metadata. The system is based on basic Dublin Core fields of bibliographic description of the document [14] to keep pace with the needs of the description of each group. The digital file of the document contains metadata and digital content and classified as combinations of the material. Thus, the construction and internal organization of the digital material and its relationship to other articles are

Table 2. Digital repositories software systems: DSpace, EPrints and Greenstone.

Characteristic	DSpace	EPrints	Greenstone
Year of creation	2002	2000	1997
Operating system	Linux or Unix, Windows, Solaris	Linux, Unix, Windows	Linux or Unix, Windows, Mac OS
Databases	Oracle, Postgre SQL	MySQL, Oracle, Postgre SQL, Cloud	Its Own
Programming language	JSP, JAVA	Perl	Perl, C++, Java
Web server	Apache, Tomcat	Apache	Apache/IIS
Source identifier	CNRI	Not applied	OAI Identifier
OAI-PMH Protocol	Yes	Yes	Yes
Storing and retrieving	We can store and retrieve all types of content	We can store and retrieve all types of content	We can store and retrieve all types of content
Metadata standard	Dublin core, Qualified Dublin core, METS	Dublin core, METS	Dublin core, Qualified Dublin core, METS, NZGLS, AGLS
Interoperability Machine—Machine	OAI-ORE, OAI-MHP, SWAP, SWORD	OAI-MHP, OAI-ORE, SWORD, SWAP, RDI	OAI-MHP, Z39.50
User interface functions	End use deposit, multi language support	End use deposit, multi language support	End use deposit, multi language support
Minimized preview	Images	Images, Audio, Video	Images, Audio, video
Search capabilities	Defined field, Boolean logic, Sorting options	Defined field, Sorting options	Defined field, Boolean logic
Browsing options	By author, subject and collection	We can browsing by using any field	We can browsing by using any field
Service delivering capability	ATOM, RSS	ATOM, RSS
User authentication	LDAP authentication, Sibolet authentication	LDAP authentication	User groups
Statistic report	Number of full record	Number of full record	Number of full record
Services	Services through third part service providers	Training Consultancy, site visit	Training

represented by structural metadata that connects each of the articles by the other to be the logical unit of those materials. DSpace uses global unique identifiers of the materials based on the CNRI Handle system. Also, permanent identifiers of digital files are used for each article. The primary entity in EPrints is data, which is a record of metadata and we can connect a single or multiple documents (files) with data entity. Each entity has unique data identifier.

- Support of collections and relations: Collections in Greenstone system are known for its several features that describe the functions of the collections. These features are; ability to indexing, searching, browsing, file formatting, and changing of plugs and access point for the import of digital content. As well as there are many features to view the collections. The representation of hierarchical structures is supports text documents for the chapters, sections and paragraphs. The definition of the specific sections in the text document is implemented through XML markup language signs. XLinks can be used in the document to link it with other documents and sources. The DSpace software supports collections of subjects or units (key collections), which consists of single collection or several collections. Also, the subject belongs to one collection or several collections, but the subject owns only one collection. It is possible to define default values for the fields of metadata in the collection. The descriptive metadata defined for the collection is the title and description, and there is no support for the relationships between different subjects. As for as EPrints software concerned, it does not have any consideration for the collections and combine data entities based on defined fields (subject, year, title, etc.). As well as there is no definition of the links between documents, except for the use of URLs in specific metadata fields.
- Digital content storage and metadata: In Greenstone software all the documents and sources stored in files system. Metadata is user-defined storage for documents and uses internal XML format. The DSpace software stores Dublin Core metadata defined in database (Oracle or PostgreSQL). The other metadata collections and digital content are represented as a digital file stored in the file system. All digital files linked to specific digital format file support all digital file formats, which shows the conservation of specific file formats. Metadata fields in EPrints are user-defined. The data entity that contains metadata is stored in database MySQL, and documents (digital content) stored in the file system.
- Interoperability features: All systems support OAI-PMH protocol for the participation of metadata for digital libraries with other

repositories. Greenstone supports Z39.50 protocol to respond to queries for specific metadata collections. The DSpace has capability to export digital entities as METS files and XML, and both the systems use permanent websites URLs to access digital content and provide unified access mechanism to external services. DSpace, also, supports open URLs protocols and provides links to each page of the subject. As for as EPrints is concerned, it issues data entities in Metadata Encoding and Transmission Standard (METS) [15] and Moving Picture Expert Group (MPEG-21) and Digital Item Declaration Language (DIDL).

- User Interfaces: In Greenstone system, default web user interface provides browsing, searching in collections and mobility within the hierarchical entities (like books) by using content schedule. To view the documents and search results it may vary depending on extensible style sheet language transformations patterns (XSLTs). In DSpace system, the default user interface on the web allows the end user to browse collections and view item specific Dublin Core metadata and mobility between the digital files and navigate inside the items supported through structural metadata that may define the arrangement of complex digital content (such as the pages of books or web pages). The user interface, makes it also possible to search by key words. In Eprints, the default user interface on web allows possibility to browse specific metadata fields that usually are (topic, subject, and date). Browsing process can take hierarchical form for subjects' fields. Also, the structure of the search allows user to define the possibility of confining search inquiry through using multiple fields and choosing evaluations form the list.
- Access control: Greenstone software's users belong to two groups of defined users. First group is the group of directors or group creators. This group is considered the first user and has the right to create, delete users, and update the group. The second group is called end users and they have right to free access to all the groups and documents. DSpace system supports users (cyber users) and groups that possess diverse rights. The identity of users is checked through the user password and certificates of Lightweight Directory Access Protocol (LDAP) and certificates of X509 designed to create a self-signed certificate file to allow access control rights for each item. These protocols determine the processes that enable user to perform these operations. These operations consist of reading, editing digital files for the item. In addition to the addition and deletion of item's collections and reading as well as editing the item. As well

as addition and deletion of the item from the group. The rights are based on default prevention policy. The users registered in EPrints system can create and edit data entities and can access by using the user name and password.

- Search and Browsing: Greenstone system provides feature of indexing text documents and defined fields of metadata. It also provides feature of search in the chapters of specified document such as title, chapter, paragraph or the entire document. The open-source applications use (gigabyte) to strengthen indexing. Browsing includes a list that enables the user to browse the selected fields using structures of hierarchical compilations. Various groups offer different browsing facilities. DSpace system provides indexing feature to basic metadata group (specified Dublin Core) by default using a relational database. In addition to indexing other specific metadata groups that are provided using Jakarta Lucene programming application interface API. Lucene supports flexible search ignores the end of words, stop, and remove the word. We can carry out restricted search in groups and units as well as Lucene provides browsing feature as a default feature. As far as Eprints system is concerned it supports indexing of each fields of metadata using database MySQL database and supports indexing the full text of specified fields. As for as integrated search of specific fields and absolute text, it available for the end user.
- Entity management: In Greenstone, new groups and collections are built by using interface of Greenstone library specialists or electronic work orders program. In DSpace system, items are created through Internet that is managed by the user and importer of item processor, which accommodates eXtensive Markup Language (XML), metadata documents and core content files. In both cases, the handling process workflow begins with formation of the group. The workflow can be configured to accommodate one to three steps and may different users or groups may interfere to provide the material. As for as the groups and units are concerned, they are created through the user interface on the web. In EPrints, default user interface on the web allows to create and edit entity. We can import entities of text files by using a variety of formats (METS, DC, MODS, BibTeX, and EndNote).
- Multiple language support: All systems use standardized coding system "UNICODE", therefore, they support multiple languages. All systems use multiple languages in the metadata fields and digital content. EPrints system provides eXtensive Markup Language (XML) feature in metadata fields to introduce the different languages that are used in metadata fields. Greenstone provides

Table 3. The various basic features of the major digital repositories software systems.

Basic features	DSpace	Greenstone	EPrints
Entity model	4	3	2
Relation and collection support	4	5	1
Metadata and digital content storage	4	3	3
Interoperability features	5	4	5
User interfaces	4	4	4
Access control	5	2	2
Browsing and searching	4	4	4
Entity management	4	2	4
Multiple language support	3	4	4
Customization level	3	4	3

interfaces ready for use in various languages that are translated in advance.

- Level of customization: Greenstone provides a customization feature based on XSLTs system to view the collection of information. As well as to view the agents who control the specific operations of the digital library. Greenstone also provides a background or internal interface containing groups as well as a front interface that is responsible for displaying collections, documents and search environment. Although DSpace has a flexible entity model, it is not very open in building a variety of entities with metadata-based groups because its database is a construction oriented method. User interface is fixed and stable and provides only offered entries. One of the flaws of this system is that it supports only specific file formats as digital content. Data entities in the EPrints, include user defined metadata and the user can write additional items to export data entities in different text formats. ACore APL is provided in programming language "Perl" for developers who prefer basis of functional availability of digital library. Through the analysis of the three software and systems for building digital repository we mentioned each basic feature of these systems according to minimum evaluation score (1) and the maximum score (5), as shown in Table 3.

6 CONCLUSION AND RECOMMENDATIONS

Every organization has multiple specific requirements that depend largely on the number of

collections, types of objects, nature of the item, multiplicity of update, distribution of content, and the time frame for the development of the repository. In each case, specialists staff members of the organization and universities must participate in the process of choosing the program to get the best result. All the programs may not be suitable for each institution. Only a single program could not suit standards of other programs as every program has features and weaknesses as we mentioned in the process of comparison. This study cannot provide a single solution but it can be used as guidelines for organizations that intend adding digital collections or send their collections to a new repository environment. From our study we recommend the following:

- It is necessary to look into the existing state of scientific and research institutions and meet their needs. In case the scientific storage of the institution is derived from scientific research and observations or survey, these materials are shared for internal or public use. The researchers can send their own collections, publish them, or choose policies to get them, or get collections of information polished by librarians or specialist staff members. The institutions have to support user registration, access policies, and link between various objects (data collections and publications). Data collections are exported in the form of common models and most of the data collection are sent in text files or tabular data. The documents usually are written documents. Thus, DSpace is the appropriate digital repository system, for example, DSpace represents units in the system (for example, departments of the university) and collections (for example, theses, and research papers).
- To meet the needs of institution that intends to publish its digital content in simple forms and for a specific time as well as prefers to integrate interfaces of the digital library with its website on internet. In this case, the most appropriate system for digital library is EPrints because it separates display from storage, and is not linked to specific metadata standards, and provides simple user interfaces to provide and display documents and metadata.
- In case the institution wishes to publish electronic books containing digital images of book's pages and files in PDF format, as well as providing content and metadata through librarians, and using digital library system for customization. Then the Greenstone is the best system because it works on the representation of books in a hierarchy manner using tabular contents while providing search feature for the full text of the document and chapters.

- DSpace digital library software is the most widely accepted among digital repositories solutions. It is comfortable functionally and supports a wide range types object types, including images, audio, text and video. It provides implementation of comprehensive guidelines.
- Greenstone and EPrints are also widely used as they are two low cost options for initial repository. They enable users to access the items and theses before printing and post-printing. As well as they allow downloading a range of object types with video, audio, images and compressed files. They also allow educational institutions to control use of these packages.
- Institutions that find EPrints programs unsuitable relatively, may find DSpace programs and Greenstone more responsive to their needs without any complications.
- DSpace supports document content policies and accommodates different types of formatting for digital documents.
- EPrints is an appropriate system for digital repositories and used on a large scale. It is also suitable when we need technical support and training for using the software.
- A number of institutions are using Greenstone and EPrints software, but the bulk of the libraries prefer DSpace software for digital repository as it has many advantages and can support many of the forms and shapes.

REFERENCES

1. Budapest open access initiative. Available on the web: <http://www.budapestopenaccessinitiative.org/read>, visited date 13/2/2016.
2. Joan M Reitz. Dictionary for Library and Information Science. Westport, Conn.: Libraries Unlimited, 2004.
3. Kifah Eisa, introduction to free software, available on: www.freesoft.jo/www/people/.../freesoftware_whitepaper_arabic.pdf, visit date: 13/2/2016
4. F Othman & F AL Zaghoul. A Comparative Study of Two Open Source Integrated Library Systems (ILS): PhpMyBiblio (version.3.5.1) and NewGenLib (version.3.0). Libraries and Information Centers in a Changing Digital Environment, Amman 29-31/10/2013, 32.
5. DSpace Federation. Available at <http://www.DSpace.org/> visited date 13/2/2016
6. Greenstone Digital Library Software. Available at <http://www.greenstone.org/> visited date 13/2/2016
7. EPrints for Digital Repositories. Available at <http://www.eprints.org/> visited date 13/2/2016
8. Amiya Kumar Das. (2015). Comparing Open Source Digital Library Software: Special Reference to DSpace, EPrint and Greenstone.- International Journal of Advanced Research in Computer Science and Software Engineering.-Vol5, No7 pp. 70-73.

9. George Pyrounakis, (2014). Mara Nikolaidou and Michael Hatzopoulos. "Building Digital Collections Using Open Source Digital Repository Software: A Comparative Study." International Journal of Digital Library Systems (IJDL). Vol 4. No1 pp. 10–24.
10. Ahamd Mahir Khafaza, (2014). Open source software for libraries and information centers: suggested standards to choose open source system for Arab libraries management. Cybrarians Journal, issue: 6 (December), available on: http://www.journal.cybrarians.org/index.php?option=com_content&view=article&id=676:opensource&catid=270:studies&Itemid=99, visit date: 13/2/2016.
11. Suha Bashir Ahmad Abdullah, (2013). Digital content management system on internet. An evaluative study to get suitable standard specification for the application of Arab information facilities. Ph.D. thesis, Binha University, Egypt, available on: http://www.eulc.edu.eg/eulc_v5/Libraries/Thesis/BrowseThesisPages.aspx?fn=ThesisPicBody&BibID=11776555&TotalNoOfRecord=4&PageNo=1&PageDirection=previous, visit date (13/2/2016)
12. Talal Nazim Al Zaheer, Aseer Majid Al Sadi, Digital repositories systems and their evaluation standards, ACADEMIA, available on: https://www.academia.edu/9943014/%D8%A7%D9%84%D9%85%D8%B3%D8%AA%D9%88%D8%AF%D8%B9%D8%A7%D8%AA_%D8%A7%D9%84%D8%B1%D9%82%D9%85%D9%8A%D8%A9_%D9%88%D9%85%D8%B9%D8%A7%D9%8A%D9%8A%D8%B1_%D8%AA%D9%82%D9%8A%D9%8A%D9%85%D9%87%D8%A7, visit date (13/2/2016)
13. The Directory of Open Access Repositories—OpenDOAR. Available on the web <http://www.opendoar.org/index.html>- visited date 13/2/2016
14. DCMI Metadata Terms. Dublin Core Metadata Initiative. Available at <http://www.dublincore.org/documents/dcmi/terms/> visited date 13/2/2016
15. METS: An Overview & Tutorial. Library of Congress. Available at <http://www.loc.gov/standards/mets/METSOverview.v2.htm>. visited date 13/2/2016

Institutional digital repositories specializing in cognitive science: An exploratory study

Fudhah Ateq Al Selami

Department of Management Information Systems, College of Business-Alkamel Branch, University of Jeddah, Jeddah, Saudi Arabia

ABSTRACT: The digital repositories are considered a new channel of scientific communication that are available on the Internet, as they are regarded as forms of the free access strategy that help in the development of intellectual output and in relaxing the restriction. The main purpose of this article is to identify the digital content of digital repositories specializing in cognitive science which applying open source software by using descriptive, analytical and comparative approach for the detection of the transition coverage of the digital content of these repositories from the temporal–spatial, numerical, objectivity and quality sides and realizing the ways of providing content, digital assets and software to build the digital repositories, its management, policies of conservation which related to the digital sources and the administration of rights according to the knowledge bank at the Ohio State University and Cogprints repository for cognitive science. The study has concluded that the groups of Cogprints repository for cognitive science still growing too slowly and that the coverage of its resources is limited to specific topics such as articles and thesis etc. Both repositories are allowing the possibility to download the full text of all sources of information to all users, free of charge and using pdf format.

Keywords: Digital repository, Cogprints repository for cognitive science, The Knowledge Bank repository at the Ohio State University and cognitive science repository

1 INTRODUCTION

Scientists gained a profit from using the digital repositories as a new channel of scientific communication that are available on the Internet, in order to create what is known today as the digital repositories, which one time were known as electronic E-print archives or Open Archives or open access repositories, they were appeared for the first time in 1991 as the first service to provide pre-release research on the internet under the name Arxiv, as the first objective digital repository dealing with physics at the global level, and was established at the hands of the physicist Paul Ginsparg. After three years it was followed by Cog-print repository in 1998 at the hands of Stephen Hernad, which led to the appearance of other repositories without interruption [1].

Digital repositories are known as “a system that store content and digital assets with the ability to save it for later search and retrieve”, so the repository has to allow the possibility of importing, identification, storing and retrieval of this asset. Repositories are considered as a kind of management systems that combine intellectual assets of an institution, and then allow using it, supporting

many of the activities within the institution [2]. ODLIS dictionary mentioned a definition of the Institutional digital repository as a set of services provided by the university for its employees for the management of their digital production and publication, the examples of this production include Thesis, studies, technical reports, batch data and teaching materials and other productions, the institution is responsible of overseeing these materials, making it available without any restriction sand, undertaking long-term saving of these materials [3].

Digital repositories allows digital intellectual production of a subject of what is called (objective) or what is related to an institution or what is called (institutional) either it was tight or loose, varied and without restriction or material condition with a minimum of legal restrictions, that will lead to provide long-term conservation of the available materials. Any other types of institutions can produces a large number of research or other intellectual production which makes it necessary to create repositories such as museums, governmental departments, international organizations, research centers, professional scientific societies and so on.

Both (Barton & Waters) indicate that the institutional digital repositories are established and used by universities and research libraries for several purposes, the most important, scientific communication, storing educational materials and curricula, posting, the management of scientific research, preservation of digital materials for long-term and improving the image of the university by showing its academic research and playing a leading role for the library at the enterprise level, knowledge management, evaluation of research and encouraging the free access to information, scientific research and host the digitized collections [4]. It also helps in research cooperation by facilitating free flow of scientific information, and increasing public perception of the research activities and efforts [5].

There are a lot of software that allow the creation of digital repositories and institutional management, including open source software, paid software and some software that are designed by a certain institution for their specific needs. Most repositories software allow creating the repositories which offer a range of functions including: registration mechanism for users, logging mechanism and downloading electronic materials with a comprehensive metadata and description, also allow reviewing data and making inquiries about before its publication, as it provides the ability to store sources and the management of their content and allow access to digital sources through the potential of internal and retrieval search, in addition to some of means that help in controlling over access to resources operation, harvest of the metadata elements by different research tools, especially search engines [6].

This study will focus on the Institutional digital repositories specializing in Cognitive Science, this include a Cogprints repository, the Knowledge Bank repository at the Ohio State University through identifying digital content of the digital repositories that applying open source software by using Descriptive, analytical and comparative approach for the detection of the Transition Coverage of the digital content of these repositories from the temporal-spatial, numerical, Objectivity and quality sides and realizing the ways of providing content, digital assets and software to build the digital repositories, its management, policies of conservation which related to the digital sources and the administration of rights.

2 PREVIOUS WORKS

The “Mapulanga Patrick study” showed the prospects and challenges which face the process of digitizing library materials and building digital repositories in five libraries of the University of Malawi. Also, the study adopted the descriptive ana-

lytical approach, they used the questionnaire as a tool for collecting data and come to the most important findings is that digitizing library and building digital repositories were applying at a snail space; however, there is an increase in frequency on digital repositories scale. The databases were victim of many viruses through the internal network and most of the technical projects lack the technical skills, especially operating systems in Linux environments. Three university libraries used Greenstone; fourth library used DSpace while the fifth one used Procite. The study recommends that, there is an imperative need to apply new strategies especially for digitizing library resources, copyright and building digital repositories in Malawi University Libraries [7].

“Suha Bashir study” dealt with the management of the digital content and its systems, also collecting and capturing Arabian and global digital projects. This study also deals with, the subject of digital content and the management of this content, as well as attempting to collect and capture Arabian and global digital projects represented in libraries, archives, repositories digital museums and other sites of open sources on different types and form son the web, and studying the numerical and spatial, linguistic and quality trends of the systems used in digital content management. Then evaluating Greenstone and DSpace as the most commonly systems that are used in the Arab world; they used survey, assessment comparative and experimental approaches. The results show the weakness of Egypt’s participation in the use of digital content management systems, and that English language came to the fore as it is used to facilitate information of the digital content all over the world [8].

The purpose of “Talal Nazim and Atheir “study is to recognize the importance of digital repositories, and their role in the conservation and organizing digital content. As well as identifying the most important international standards of assessing the creation of digital repositories system for the purposes of comparing them in order to give an opportunity for institutions to choose the appropriate information [9].

Brown & Abbas study dealt with the emergence of digital repositories in the United States, the study focused on specialized repositories in biology (indicating its role in the research process at the University of Oklahoma, by conducting personal interviews with twenty scientist who are specialized in geosciences and biology about their ability to allow their research on the Internet for the Exchange of information and knowledge. the study showed that a few of them are interested in creating digital repositories, and most of them does not have enough time to create or manage the digital repositories [10].

Ahmed Abaad El-Araby study aimed to examine the fiftieth repositories and analyzed them according to the classification of The Cyber metrics Lab, distributing its contents numerically, qualitative, chronologically and Objectivity, and determining methods of search and retrieval, software used and policies, to develop a mechanism for the establishment of repositories to guide the Arab universities when building a digital repositories. The study has adopted the analytic, descriptive approach and the authors found that the repositories are using various methods that enable users of retrieving different forms of information, 75% of the repositories used open source software, Eprints program was one of the most frequently used programs by 45.8%.

These repositories have developed their own policies at the virtue of 45.8%. We must refer that most of the previous studies focused on providing criteria to evaluate digital repositories and trying to identify the digital content to the repositories of global institutions for the detection of the Transition Coverage of the digital content of these repositories from the temporal-spatial, numerical, Objectivity and quality sides and realizing the ways of providing content, digital assets and software to build the digital repositories, its management, policies of conservation which related to the digital sources and the administration of rights at the Knowledge Bank repository at the Ohio State University and Cogprints repository for cognitive Sciences.

3 PROBLEM FORMULATION

The digital repositories in general and specialized institutional repositories in particular are considered the core of research and study, where the researcher is trying to shed the light on digital repositories specializing in cognitive science to the institutional repositories of open source software through the following questions:-

- What are the spatial and temporal, numerical, objectivity and quality of institutional digital content in repositories that are specialized in the field of cognitive science?
- What are the methods of providing content and digital assets of the institutional digital repositories that are specialized in the field of cognitive science?
- What are the policies that used to save digital resources and rights management in the institutional digital repositories that are specialized in the field of cognitive science?

4 RESEARCH METHODOLOGY

This study has adopted the descriptive, analytical and comparative approach for studying the

institutional digital repositories specializing in cognitive science which represented in Knowledge Bank repository at the Ohio State University and Cogprints repository for cognitive Sciences to identify the nature of the content and digital assets of those repositories. Also we study the subject in database, and web sites related to the subject of the study to draw out a checklist of basic characteristics of digital content repositories. Then the researcher examines the repositories, collecting information for the preparation of the sources of information, forms, types and the potential of search, browsing and policies used and other aspects covered by the study.

5 MAIN FINDINGS AND RESULTS

Digital repositories systems rely on the deposit of digital sources by registering function which defines ways to send personal scientific sources of researchers or employees, as well as, establish in repositories in line with international standards as the Initiative of Open Archives to ensure search in the repository content by engines and search tools, informing the beneficiaries of new materials regarding their interest by using really simple syndication service or setting up list of new updates. Digital repositories depend on long-term achieving which deposited by workers who are affiliated to the foundation by applying the best generally accepted standards and conservation programs. (See table No 1).

Comparisons of the digital repositories are done according to the basic characteristics mentioned below:

- Temporal coverage: Knowledge Bank repository at the Ohio State University covers the digital entities from 1920 to 2016 while Cogprints repository of cognitive science covers only from the year of 0013 to 2015.
- Objective coverage: Knowledge Bank repository at the Ohio State University covers all areas because the entities are arranged according to the units that comply with the administrative units of the University of Ohio such as schools, departments and research centers and labs, while Cogprints repository of cognitive science covers only the fields of psychology, neuroscience, linguistics, computer science and Philosophy such as mind, language, knowledge, science, logic, biology, medicine, anthropology such as high-end zoology Meta cognition and Learning, archeology, social sciences, Mathematical sciences and other forms of science which are related to cognition study.
- Numerical coverage: Knowledge Bank repository at the Ohio State University contains about

Table 1. The most prominent technical and functional potentials of the digital Repositories specializing in cognitive science.

Characteristic	Ohio state university knowledge bank	Cogprints cognitive sciences archive
The year of establishment	2002	1998
designation of origin or university	Ohio State University	University of Southampton
geographical position	United states of America	United kingdom
website	https://kb.osu.edu/dspace/	http://Cogprints.org
Open source software	DSpace	Eprints
logging	Available	Available
Number of digital entities	70638	21522
Sources indicator	CNRI	None
OAI-PMH protocol	Yes	Yes
Storing and retrieval	It can restore and retrieve all forms of resources	It can restore and retrieve all forms of resources
Metadata standard	Dublin core, specific subset of the Dublin Metadata Encoding and Transmission (METS)	Dublin core and METS
virtual machine interoperability	OAI-MHP, OAI-ORE, SWORD, SWAP	OAI-MHP, OAI-ORE, SWORD, SWAP, RDI
User interface functions	End-user deposit	End-user deposit
User interface language	English	English
Thumbnail preview	Images	Audio, video and images
search capabilities	Selected field, Boolean logic and sort options	Selected field and sort options
Browsing options	Author, subject and collection	Browsing by using any field
service delivery	RSS	RSS ATOM

70638 digital entities, while Cogprints repository of cognitive science contains 21522 digital entities.

- Quality coverage: Knowledge Bank repository at the Ohio State University contains articles, thesis, books, book chapters, research papers, order of business, (unpublished) drafts, presentations, posters, conferences fact sheets, multimedia and University publication, while Cogprints repository of cognitive science contains articles, thesis books, book chapters, research sheets, conferences fact sheets and (unpublished) reports and drafts.
- Free access to sources of information in repositories: both repositories allow downloading the full text of all sources of information to all users free of charge, using pdf form at file.
- Methods of search and retrieval in repositories: the repositories allowed several techniques that enable its users to retrieve various information sources, including sources retrieval by browsing and retrieval of resources by searching in the Internal search engine:-

- First: retrieving the resources by browsing: default user interface helps in the retrieval of the sources in Knowledge Bank repository at the Ohio State University by browsing through units that contain one group or several groups that were established by faculties and departments of the University, the title, author, subject, date, and displaying Dublin core, specific subset of the Dublin Metadata of the material, as it allow the collective mobility among digital files and supported materials through the structural metadata that may determine the order of the complex digital content(such as pages of books or web pages). The default user interface also allows searching with keywords. While the user interface on Web at Cogprints repository of cognitive science, retrieving sources by browsing the subject and history and offers the possibility of browsing metadata fields. Browsing process takes hierarchical form. The entities can also be imported through the text files by using (METS, DC, MODS, BibTeX, End-Note) formats.

- Secondly: retrieving the resources by searching the internal search engine: both repositories allow the Advanced Search service which allows searching depending on certain time, or on behalf of the author, the subject or title, as well as the Boolean Search. As for other search potentials of Knowledge Bank repository at the Ohio State University such as searching by words, author, title, extractor, type of material and its serial number. The other research possibilities of Cogprints repository of cognitive science are represented in searching by words, author, title, full text, extractor, type of material, the Scientific Section and the state of art and history, depending on the published material either it was published or In Press material.
- Searching the internet: the Knowledge Bank repository at the Ohio State University allows searching in the contents of college and specialized institutions, specialized evidence and research groups as it allow searching Google search engine through a search icon of Ohio State, while Cogprints repository of cognitive science does not allow this.
- Descriptions of sources of information in repository: they display the citation of the resource that includes all bibliographic database needed to the citation of the source, then displaying a set of detailed source data. Description data of the Knowledge Bank repository at the Ohio State are represented in title, author, date, publisher, link, subject, the type of material and extract. As for Cogprints repository of cognitive science, it allows only the Title, Author and history.
- Ways of displaying data description of sources of information in repositories: ways of viewing the descriptions of data are varied, as the descriptions of data in the Knowledge Bank repository at the Ohio State are displayed according to Dublin core plan, while Cogprints repository of cognitive science displays data according to more than one descriptive plan such as Dublin Core, metadata object description schema(MODS), MA chine-Readable Cataloging (MARC) and other plans of Metadata, as it displays data in different forms such as Endnote, Reference Manager, Text, Excel, ASCII Citation.
- The arrangement of recovered results: the results in the Knowledge Bank repository at the Ohio State are arranged, according to the link, the history in ascending or descending order, the listing date and title. While in Cogprints repository of cognitive science they arrange the recovered results on the date of publication in ascending or descending order, the title and the author.
- Additional services of repositories: repositories allow a group of services as well as providing search and free access to the sources. The

Knowledge Bank repository at the Ohio State enables Really Simple Syndication service, user preferences user statistics and feedback. Cogprints repository of cognitive science allows only Really Simple Syndication service.

- Policies of repositories: they are the policies which determine the available resources and the necessary steps for listing them. Also determining the types of files that will be inserted, simply they are the listing policy, Quality Assurance, management of the digital content. Both repositories paid no attention to quality control policy, content management, copyright policy and the policy of conservation. The Knowledge Bank repository at the Ohio State mentioned the listing of the digital entities, while Cogprints repository of cognitive science never mentioned this.

6 CONCLUSIONS AND RECOMMENDATIONS

Institutional digital repositories specializing in Cognitive Science are considered as a source of providing intellectual production to researchers in all fields of knowledge. The Knowledge Bank repository at the Ohio State and Cogprints repository of cognitive science are two repositories that allow full text digitizing digital content without restrictions; however, there are differences in the coverage of the digital content and methods of providing digital assets, software of digital repositories and management and conservation policies which are as follows:

- However that Cogprints repository of cognitive science is the second global repository in construction, but the collections are still growing very slowly. In 2015 they deposited only three entities. The Knowledge Bank repository at the Ohio State prove superior in the number of deposited entities which reached about 70 638 Digital entity.
- The objective coverage of The Knowledge Bank repository at the Ohio State include all areas because it covers all the substantive disciplines of the University, while the Cogprints repository of cognitive science covers a relatively specific fields.
- The quality of coverage of the sources are similar in both repositories, in the ability to deposit articles, thesis books and chapters of books, research papers, order of business, (unpublished) reports and drafts, presentations, posters, facts Conference, multimedia and publications of the University.
- Both repositories allow the possibility to download the full text of all sources of information to all users free of charge and in pdf form at file.

- The repositories allowed several methods that enable users to retrieve various information sources including the recovery of resources by browsing, where The Knowledge Bank repository at the Ohio State has multiple browsing systems through units, the title, author, subject and date. The Metadata are displaying in specific subset of the Dublin Metadata, while in Cogprints repository of cognitive science resources are retrieved by browsing the subject, history and offers the possibility of browsing metadata fields. Browsing process takes hierarchical form. The entities can also be imported through the text files by using (METS, DC, MODS, BibTeX, EndNote) formats.
- Both repositories enable the Advanced Search which allow searching by a specific time or by the name of author subject or title, as well as searching by Boolean Search.
- Cogprints repository of cognitive science lacks to allow searching the internet On the contrary The Knowledge Bank repository at the Ohio State that allow searching libraries content, specialized institutions and research group sand the ability to search Google search engine.
- The Knowledge Bank repository at the Ohio State provides Really Simple Syndication service, user preferences and the feedback of the repository, but Cogprints repository of cognitive science allows only Really Simple Syndication service.
- Both repositories need to determine conservation policy of repositories, quality control, and management policy of the digital content and copyright management. The Knowledge Bank repository at the Ohio State allows the inclusion of digital entities policy.
- The Institutional digital repositories specializing in Cognitive Science must set up policies for inclusion of sources and intellectual works of repositories, quality control, the content management and management of copyright and conservation in the two repositories.
- The repositories must make a Opinion survey and interviews with beneficiaries to support the effectiveness of the repository.

REFERENCES

1. Pauline Coisy. Archive ouvertes, HAL, HAL-UPMC. BUPMC. 27/05/2011. P.04. Available on the web: http://www.jubil.upmc.fr/modules/resources/download/bupmc/docs-bu/8_HAL-UPMC/images/Presentation_BUPMC.pdf visited date 3/3/2016.
2. Hayes, H. (2005). Digital Repositories: Helping universities and colleges. Available on the web [http://www.jisc.ac.uk/uploaded_documents/JISC-BP-Repository\(HE\)-v1-final.pdf](http://www.jisc.ac.uk/uploaded_documents/JISC-BP-Repository(HE)-v1-final.pdf)—visited date 3/3/2016.
3. Joan Retiz, (2010). ODLIS Online Dictionary for Library and Information Science Available on the web Available at: [-http://www.abc-clio.com/ODLIS/odlis_i.aspx](http://www.abc-clio.com/ODLIS/odlis_i.aspx)—visited date 3/3/2016.
4. Mary R. Barton, & Margaret M. Waters, (2005). creating an Institutional Repository: LEADIRS Workbook. Cambridge, MA: MIT, 2004. P11. Available on the web: <http://dspace.mit.edu/handle/1721.1/26698>—visited date 3/3/2016.
5. The Repositories Support Project (RSP). Available on the web: <http://www.rsp.ac.uk/start/before-you-start/benefits>—visited date 3/3/2016.
6. Ahmed Ebada El-Araby, (2012). The digital repositories of academic institutions, and their role in developing the educational process, research and drawing up a mechanism to create a digital repository of Arab universities. King Fahd National Library Journal, GS 1, 18mg.on the following link:http://www.kfnl.org.sa/idarat/KFNL_JOURNAL/m18-1/pdf/alaraby.pdf6
7. Mapulanga Patrick, (2013). “Digitizing library resources and building digital repositories in the University of Malawi Libraries”, The Electronic Library, Vol. 31 Iss: 5, pp. 635–647.
8. SuhaBashir Ahmed Abdel Aal. (2013). Digital content management systems that are available on the World Wide Web (Web):An evaluation study to create standards suitable for applications of Arab information facilities, Ph.D. thesis, Banha University, Egypt-. Available on http://www.eulc.edu.eg/eulc_v5/Libraries/Thesis/BrowseThesisPages.aspx?fn=ThesisPicBody&BibID=11776555&TotalNoOfRecord=4&PageNo=1&PageDirection=previous reviewing date: 2/13/2016.
9. NazimTalalAl Zuhair, Athermajed El-sady: digital repositories systems and standards of evaluating available on https://www.academia.edu/9943014/%D8%A7%D9%84%D9%85%D8%B3%D8%AA%D9%88%D8%AF%D8%B9%D8%A7%D8%AA_%D8%A7%D9%84%D8%B1%D9%82%D9%85%D9%8A%D8%A9_%D9%88%D9%85%D8%B9%D8%A7%D9%8A%D9%8A%D8%B1_%D8%AA%D9%82%D9%8A%D9%8A%D9%85%D9%87%D8%A7 reviewing date: 2/13/2016.
10. Brown, C. & Abbas, J. M. (2010). Institutional Digital Repositories for Science and Technology: A View from the Laboratory. Journal of Library Administration, 3, pp. 81–215.
11. Ahmed Ebada El-Araby, (2012). The digital repositories of academic institutions, and their role in developing the educational process, research and drawing up a mechanism to create a digital repository of Arab universities. King Fahd National Library Journal, GS 1, 18mg.on the following link:http://www.kfnl.org.sa/idarat/KFNL_JOURNAL/m18-1/pdf/alaraby.pdf6

Enterprise architecture moving from professional certificates into academic credentials

Fekry Fouad

King Abdul Aziz University, Jeddah, Saudi Arabia

ABSTRACT: EA “Enterprise Architecture” considered as a best practice, for a multi-disciplinary, workflow enables to plan for a smooth transformation of an organization using a strategic analysis, also EA is a professional work and proceeding of a practice to assist enterprises in solution designing to achieve the future business objectives. Today’s the rapid-changing of business operating models and technology options are evolving at an ever-increasing rate, and EA enables organizations to move up of the curve. However, what does the future really hold for Enterprise Architecture as Academic Credentials. It is clear that there is a major difference between academic credential and professional practice of EA. The professional practices are ‘certain course of actions’ prove the competence’ or ‘power of expertise’ whereas academic credentials almost not. Certification is only a prove to make its holder more valuable as a worker or employee. It immediately recognizes evidence and skill sets credibility, when trying to have a new vacancy or job and considered as an advantage in career advancement.

This paper details the Enterprise Architecture to move from Professional Certificates into Academic credentials” to provide an opportunity for education researchers, practitioners, professional bodies of Enterprise Architecture, as well as IT educational Academy Staff from around the globe to exchange their experiences, ideas, theories, strategies, and technology-inspired solutions for achieving more engaging, more efficient, more accessible, and more successful IT education globally. The paper suggest a curriculum for a good combination of both theoretical knowledge and practical understanding and the methodology of moving from Professional Certificates into Academic credentials for Enterprise Architecture”.

Keywords: EA, Enterprise Architecture, Professional Certificates, Academic Credential

1 INTRODUCTION

“Credentials” always refer to a specific qualifications education or academic outcome, for example a completed diplomas or degrees that somebody have completed or even partially “Credentials” can also refer to occupational qualifications, such as professional certificates or work experience, [1].

The Enterprise Architecture professional or occupational credential are stabled enough through different professional entities such as an academic degree or professional designation such as PhD, P.Eng or M.D., whether this be purely honorary or symbolic, or associated with credentials attesting to specific competence, learning, or skills [2].

The main paper objectives is to give a clear vision for the Enterprise Architecture academic credential degrees such as College diploma, Bachelor’s degree, Master’s degree, PhD or Doctorate degree, inline with the professional entities work in the filed of Enterprise Architecture.

The designations of the academic credentials and its related experience required for the Academic Staff will be in high need to make sure for them to

have a highest quality of education for the know how of the enterprise architecture to meet the relevant accreditation standards—if any-.

The standard minimum qualification for teaching in an EA bachelor degree program will be a Master’s degree in the discipline of EA or related field. The College desires to have Instructional Staff in degree programs that hold the terminal degree in the EA discipline, which often would be a PhD. While the professional credential or certificate Occupational or Professional Credentials are given by a professional bodies, and specially for the enterprise architecture such as the Federation of Enterprise Architecture Professional Organizations (FEAPO) [3], that is a worldwide association of professional organizations which have come together to provide a forum to standardize, professionalize, and otherwise advance the discipline of enterprise architecture. The professional credential that can be given by the FEAPO through its affiliated bodies are:

Professional licenses, Memberships in professional associations, Apprenticeships, Trade certificates, Work experience accredited certificate.

The paper show the best proper curriculum of the EA discipline that alternatively a combinations of Academic Credentials, occupational/professional designations and experience qualifications combined with demonstrated and documented professional competence in the field of Enterprise Architecture.

2 THE APPLICATION AREA AND NORMATIVE REFERENCES

This curriculum of scientific practical-oriented course “Enterprise Architecture Modelling (EAM)” declares the minimum requirements to the knowledge and a capability of students obtained after hearing to the course content. It determines the content of classes and type of knowledge control (examination).

This curriculum is intended for tutors that are running the EAM as an academic syllabus. The syllabus was developed in correspondence with the best practice of using EA in the real business careers.

3 LEARNING GOALS

The course «Enterprise Architecture Modelling» is aimed on acquiring the advanced level of knowledge about the Enterprise Architecture and business modelling. This course provides models and mechanisms of creation of IT architectures that are used in Russian industry companies. Two cases are aimed on practical implementation of the knowledge acquired.

4 THE COMPETENCES OBTAINED DURING THE STUDIES

After passing the exam of MEA students should:

Know:

- methods and tools of Enterprise architecture modelling;
- methods of EA optimization;
- benchmarks for B/IT efficiency;
- business functions and system of business management;
- methods of functional business-tasks analysis.

Have an ability:

- to manage enterprise architecture
- to manage life-cycles of Information Systems (IS)
- to use methods and instruments for adjusting business models;
- to use methods of innovational and entrepreneurial management.

This course is aimed on acquiring the following competences:

- Understanding the increased role of IT for companies’ business.
- Newest approaches for EA modeling.
- Application of advanced architecture principles, models and standards.
- Methodologies of EA modeling.
- Understanding of change management methodology.
- Acquiring practical knowledge for EA constructing.

Studying this course student obtains the following competences:

Competence	Descriptors—main features of knowledge acquiring (indicators of result achievement)	Forms and methods of learning that contribute to formation and development of competence
Ability to work with information from different sources	Student uses all available information sources	Seminars, Case-studies
Enterprise Architecture analysis	Student shows possibilities of selecting types of EA	Based on both: Academic rules and professional practice credentials
Capability to formulate scientific goals in EA research	Student shows the understanding of the practical field of EA	
Application of program components for handling, analysis and classification of information	Student shows the application of necessary instruments for researchers	
Preparation of scientific (technical) reports, presentations, scientific publications	Student uses reports, presentations during seminars and prepares publications on the topic	
Develop and integrate the Enterprise architecture components	Student analyses the possible ways of EA, provides arguments for and against of application of IA type, shows understanding of architecture development process	

5 PLACE OF THIS COURSE IN THE EDUCATIONAL PROGRAM STRUCTURE

At all Universities or Higher School of Economics or Management or Information Technology disciplines this course is taught for master students on Business informatics and master students on Big Data Systems. This course is based on the following subjects in bachelor program:

- Scientific seminar “Enterprise architecture”.
- Bachelor course “System analysis”.
- Bachelor course “The foundations of business process modelling”.

6 DURATION OF COURSE AND TYPES OF EDUCATIONAL WORK

Type of work	All hours/ points	Module			
		1	2	3	4
Classes (all)		–	32	–	–
inc:	–	–	–	–	–
Practical Tasks (PT)-1	–	–	16	–	–
Practical tasks (PT)-2	–	–	–	–	–
Seminars (S)	–	–	16	–	–
Experimental classes (EC)	–	–	–	–	–
Personal preparation (all)					
Writing assignments	–	–	1	–	–
Graphical work	–	–	–	–	–
Abstracts	–	–	–	–	–
Home assignments	2	–	2	–	–
Exam	1	–	1	–	–
Overall amount of hours	32	–	32	–	–
ECTS	3	–	X	–	–

6.1 Content of topics

Topic 1. *Strategic management* of enterprise from the viewpoint of enterprise architecture (theory of management: (Drucker, Chandler, Zeltsnik, Minzberg, Aacker, Ansoff, Adizes, Traut.); interrelationships of key managerial theories and concepts in area of *strategic management* with modern methodologies of enterprise architecture) [4].

Topic 2. *Operational enterprise management* from the viewpoint of enterprise architecture (Drucker, Minzberg, Norton and Kaplan (BSC, ABC). Organizational theory, lean manufacturing, 6 Sigma, Kaidzen, TOC, ISO9000/ISO20000/ISO27000, TBL and others.); interrelationship of

key managerial and concepts in area of *operational management* with modern methodologies of enterprise architecture). [5]

Topic 3. Information and socio-technical systems from the viewpoint of enterprise architecture (interrelationship of enterprise architecture with standards and practices of system and software engineering ISO15288, ISO15926, RUP, RM-ODP, Agile; with standards and practices of project management PMBoK, IPMA, P2M, MSP, PRINCE2), [6].

Topic 4. Foundations of financial management for enterprise (foundation of budgeting, financial and managerial accounting of IT asserts, liquidity, cost/benefit analysis, shareholders earnings, structure of capital, optimization of IT asserts budgeting, peculiarities of Information Systems in accounting, construction of financial—information asserts interrelation structures with EA methodologies, *financial architecture* of enterprise). [7]

Topic 5. Review from industry about popular EA methodologies: Zachmann, META Group, DoDAF, FEA, LEAD и TOGAF. (EA development, Zachmann framework, structure and model of architectural description, META Group, LEAD, DoDAF, FEA, TOGAF; other methodologies: choice of the best in the particular case). [8]

Topic 6. EA modelling according to DoDAF. Modelling of capabilities and standards (Capability view, Standards view).

Topic 7. EA modelling according to TOGAF. Instruments and ways of modeling. (Architecture Development Method (0–7)): Preliminary phase, Architecture vision, business architecture, Information systems architecture, Technology architecture, Opportunities and Solutions, Migration planning, Implementation Governance).

Topic 8. EA modelling according to TOGAF. Instruments and ways of modeling. Architecture change management, Requirements management. Enterprise continuum, ACF and other parts of methodology.

Topic 9. Peculiarities of EA modeling of large scale holding structures: architectural patterns for appropriate scaling.

Topic 10. Service-Oriented Architecture (SOA) and its relevance in modern EA. Future concept development, alignment of business processes with «cloud» solutions.

Topic 11. Practical cases of EA modelling (telecommunication industry, cement, oil & gas industry).

Topic 12. Practical cases of EA modelling (higher education, banking industry, etc.)

6.2 Course parts and types of classes

No. I/II	Name of the topic	Classes (Hours)			Home assignments	Exam	Hours
		Practice	Seminars	All			
1	Business and information technologies	1		1		1	
2	Best IT-practices and application areas	1		1			
3	Archimate modelling language	2	2	4			
4	Development of enterprise architecture against TOGAF methodology	10	6	16			
5	SOA and cloud solutions in Enterprise Architecture	1	4	5			
6	Practical cases	1	4	5			
	Sum	16	16	32			

7 METHODOLOGICAL RECOMMENDATIONS FOR THE COURSE

7.1 Form of exam and the final mark structure

Exam = 50%, 50% = essays, cases, seminar activity.

7.2 Similar questions and tasks for exam

Topic 1.

1. Enterprise architecture definition.
2. What are the parts of EA? What is their purpose?
3. What is the role of IT in business?
4. What is the role of IT strategy and IT-architecture in business changes?

Topic 2.

1. What is Balanced Scorecard and how it is used for strategic governance?
2. What is EFQM Excellence Model?
3. What is ISO 9001:2000/ISO 20000/ISO 27000?
4. What is COBIT?
5. What is ITIL?
6. What is CMMI?

Topic 3.

1. How many rows and columns does Archimate model have? Why?
2. What elements in Archi are used to model the business architecture?

Topic 4.

Which elements are used for information architecture and technological architecture modelling in Archi?

Topic 5.

1. Please name the structure of Architecture description of Zachmann model?

2. Which other models you know?
3. Which architecture methodology you suppose is the best?

Topic 6.

1. Name the main phases of ADM TOGAF.
2. What are the main outcomes of Preliminary phase?
3. What are the main outcomes of Architecture vision?
4. What are the main outcomes of Business architecture phase?
5. What are the main outcomes of Information systems architecture?
6. From what parts Information Systems architecture consists?
7. Describe the concept of application architecture and data architecture
8. Describe the concept of integration architecture.

Topic 7.

1. Define technology architecture.
2. What are the main outcomes of Technology architecture phase?
3. What are the main outcomes of Opportunities and Solutions phase?
4. What are the main outcomes of Migration planning phase?
5. What are the main outcomes of Implementation Governance phase?

Topic 8.

1. What are the main outcomes of Architecture change management phase?
2. What are the main outcomes of Requirements management phase?
3. What for TOGAF Repository is used?
4. What TOGAF extensions do you know and how they are used?

Topic 9.

1. Describe the peculiarities of SOA.
2. Describe connection of SOA notion with Enterprise architecture.
3. What are the advantages of cloud solutions whilst IT architecture development?

8 CONCLUSION

In the field of Enterprise Architecture, the paper tried to assist companies, universities, government agencies, and other employers to combine the academic and professional credentials of potential new concept of EA practicing. This research provides extensive, curriculum intended to whom it may concern in the field of EA to join the academic and the professional credential in on concept of curriculums and always review and compare and prepare the proper assessment by the academic staff for the educational track record to determine any gaps in educational progression, and whether anything is missing.

This paper is just a start for joining the academic credential with the professional credentials, the next step is a book chapter for the same subject.

REFERENCES

1. "Terminology Documents". Institute for Credentialing Excellence. Retrieved 2012-08-02.
2. "Helping Low-Income Adults and Disadvantaged Youth Earn Credentials and Build Careers: Leading Foundations Speak about Policy Priorities". Center for Law and Social Policy. Retrieved 2011-08-09.
3. Stephen IBaraki. "Brian Cameron: Professor and Executive Director, Center for Enterprise Architecture, Penn State, Founder FEAPO." at stephenibaraki.com, 2011. Accessed 24-03-2015.
4. Schekkerman, J. (2008). Enterprise Architecture Good Practices Guide—Chapters 6–8. Victoria, BC, Canada: Trafford Publishing.
5. Hanschke, Inge. (2010). Strategic IT Management, Chapter 4. Berlin Heidelberg: Springer-Verlag.
6. Abrahamsson, P., Salo, O., Ronkainen, J., & Warsta, J. (2002). Agile Software Development Methods: Review and Analysis (No. VTT 478). Oulou, Finland: VTT Technical Research Centre of Finland.
7. Avizienis, A., Laprie, J.-C., & Randell, B. (2004). Basic concepts and taxonomy of dependable and secure computing. *IEEE Transactions on Dependable and Secure Computing*, 1(1), 11–33.
8. Eason, K. (2007). Local sociotechnical system development in the NHS national programme for information technology. *Journal of Information Technology*, 22(3), 257–264.
9. Business Architecture: A Practical Guide by Jonathan Whelan and Graham Meaden. Gower Pub Co (August 28, 2012), 271 c.
10. Kamennova, M.S., Gromoff, A.I., Ferapontov, M.M., & Shmatyuk, A.E. Business modelling. ARIS technology. – M.:, 2001. C. 36–115.
11. Know Service-Oriented Architecture (SOA): Concepts, Technology, and Design, Prentice Hall (August 12, 2005), c. 792.
12. Enterprise architecture. [Electronic resource]. URL <https://learn.open2study.com/mod/youtube/view.php?id=42933>.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Digital warehouse for research in economics: Analytical study between reality and expectations

Ebtesam Hussain ALZahrani

Department of Information Science, King Abdul-Aziz University, Jeddah, Saudi Arabia

ABSTRACT: In the last days, the digital warehouse appeared with its different types, for the first times its numbers increased and its usage also.

Today, digital repositories become one of the most recent institutions of digital information on the internet with this clear increase, specialized repositories in specific fields appeared such as: economic, languages, and computer. For these reasons this research focused on this kind of specific warehouse and present comprehensive study about digital warehouse in searches of economics. Research starts with systematic introduction, then it presents digital warehouse of economy, it presents hard wares, soft ware's, human competencies, size of digital warehouse, and the way of searching on warehouse/how to search on warehouse, in addition to availability and published searches. The search focused on measurement and research on what mentioned before by focusing on different studies which review digital warehouse and its essence, generally. It also focused on digital warehouse which specialized in economy to know the importance of it in the universal universities and scientific institutions. The search aimed to present a conception of the specialized digital warehouse now/today to clarify similarities and differences between reality and what hopped from it by analysis points of strengths and weakness, and show threats and chances according to the SWOT analysis.

According to what mentioned before, the most prominent results, which researcher arrives to it, ensure the digital warehouse of researching on economy (REPEC) work to help people to arrive to the biggest collection of economic researches on the internet, newspapers and reviews. it also save the easiest ways to arrive to content and benefit from it. The (REOEC) has admen chick information and save instructions.

Keywords: Digital repositories, (RePEc) Research Papers in Economics, digital libraries, scientific and research institutions, institutional digital repositories and open access

1 INTRODUCTION

Scientific and technological development and communication Revolution led to double knowledge for the first time. In a short period access to knowledge became one of the most important things, now days. Administrators and workers at educational institutions forced to make efforts to get along with the modern, followed methods and strategies to make educational system in universities and others educational environments get along with knowledge societies. Despite the appearance of electronic education and its applications the increasing of information and allowing it randomly, this lead to interference among terminology in most of knowledge fields.

As mentioned, it is necessary to find a way to save different educational subjects, sources of information and reduce the cost of contents repetition. Hence, appeared the importance of digital repositories in educational environments generally,

and in academic one particularly. It is considered a new vision, that helps to create useful experiences and practices, as a result to the reuse of educational contents which saved on these digital repositories. Development does not stop to create digital repositories but this technological development extends to make digital repositories in specific fields to make the benefit of this information that is in digital repositories in the highest level such as now days.

So, this study aimed at focusing and careful analysis to digital repository for researching paper in economics to cover/present theses, research cuttings, magazines, and a series of papers of working which centered on searches of economics in depth and particular; to know the structure, contents, techniques, and tools of digital repository in economics, also the accounting policy/strategy which followed in creativity, retrieval, or availability; in addition to know the points of strength and support it in the other hand the points of weakness and

remove it. The researcher used the case study and collected data depends on a list of previous studies and what are included such the written statics of the repository to assurance quality and dependence. Hence the researcher show the problem of study in one basic point that is to what extent the research repository in economics achieved what is expected from; through the researcher put an analytical study which depends on/based on a collection of assumptions that she tries to achieve of it through the parts of the search, to show the relation between reality and expectations of digital repositories in economics.

2 RELATED WORKS

There is no doubt digital repositories became one of the most important standers to evaluate researching and scientific institutions, so there is global direction to make institutional digital repositories in the institutions of high education. The increasing of this education led to that it is necessary to make digital repositories with more specialized to assure complete benefit of it. Such as the digital repository in researches and economics REPEC which noticed that it serve a lot of aims, and positively helps on up grad of the quality of scientific searches and educational process generally.

Through the viewing of intellectual production we arrived to some studies that has related to our current research, we will overview the previous studies as follows:

“The first (1) study aimed to present directions of authors and their usage of institutional digital repositories as the study checks authors’ directions and ceremony of publishing in Canfield university and their worries about this, their awareness and usage of institutional repositories such the QUEprints repository, also to know the most prominent available chances to repositories in order to assurance the quality of the content of electronic learning, and the most prominent challenges that faced it’ the point of view of the professors in Saudi universities as they are one of substrates of quality of electronic education. The study reached to many authors did not hear about the QUEprints repository or its aims/targets. Authors showed/clarified although it is important to put a copy of their searches in the repository many of them do not know how to put it they depend on library. Also they have some worries about pitting their works in the QUEprints repository and it is an additional work with work pressure. Researchers recommended the need to expand in establishing and development of digital repositories.

The second study (2) aimed to describe establishment of digital repository includes learning

development units, settings of electronic tests, and design banks of questions to the students at faculty of education, Mansoura university (in Egypt). The searcher followed the semi—empirical approach and description analytical method to define the digital entities and its important, types, and targets/aims, organize and retrieval digital entities in digital repositories on the internet. The role of digital repositories to save and retrieval digital entities, and know sages of establish and testing the digital repository which founding of information and libraries department at faculty of arts, Menoufia university (in Egypt). The study reached to there are many types, and shapes of digital entities, they include digital repositories which regard one of the most important storing places of digital entities, and doubled to Arabic assistance on the map of open access. The Dspase is one of the best and famous programs of managing institutional digital repositories on the internet. Most of repositories depend on double corn meta data stander. The study recommends the importance of every university establish digital repository to publish searches of professors and increase the Egyptian assistance on the internet to courage the culture of open access on through seminars and conferences, and digital repositories’ awareness of the important of digital deposition.

The third study (3) aimed to know the reality of digital repositories and encourage professors to put their scientific production by the use of archiving resume. The searcher uses the survey method to know the reality of Arabic repositories to analysis it and deduce some of general indicators and develop them. The study reached to there is weakness in strategies which clarify deposition process, there are 50% of Arabic repositories allow/present subjects as tow and half summery. It also reached to English language go top of used files in repositories. The researcher recommended encourage professors to complete deposition, also she ensure facilitate personal deposition through website and Arabic reposition connected together.

The forth study (6) aimed to build/make suggested model within quality standers and its effects on some learning sides of students at faculty of education. The researcher determines quality standers of digital repository and its effects on some sides of learning (collection, thinking, innovation, tendency). The sample of study consists of thirteen student of third division at computer teacher department in faculty of specific education, Kafr El sheikh university (in Egypt), who have computers connected to the internet. The researcher used experimental method, survey method, and analytical order method. The study reached to there are differences in averages degrees of experimental group in both before and after

application in knowledge collection, test the ability to think and innovation, measure direction to the use of digital repository in learning for the sake of after application.

The fifth study (8) aimed to evaluate opened digital repositories on the web, reveals to what extent Egyptians researcher benefit of opened digital repositories, in addition to putting assumption of opened academic digital repository on web. The most prominent results the study reached to are the efforts to the open access, in the Arab world, are individual and very slow, there is no institutions support it. Donations get on first rank/place as one of sources of current support to digital repositories, the subject of study, while the budget of library is a fixed source of future support. The internet and associates are the tow primary sources of sample of study surrounds of open access and know open digital repositories, the website is the first source of librarians' surrounding of open access and digital repositories as one of his purview.

The eighth study (9) is the role of digital repositories in educational process in Oklahoma university in USA, it also studies the establishment of digital repositories in USA, also it focused on the repositories that specialized on biological sciences. Researchers used interview as a means of study hence the sample consists of 20 researchers specialized on anthropology and biology. The study took/presented sample's acceptance of the publishing their searches on the internet to exchange knowledge, they used experimental method. The study showed that there is an important role of digital repositories in educational process, it also clarified that little of the sample want to establish digital repositories and most of them have no time to establish and manage them.

The seventh study (7) and last study it aimed to evaluate the repository of Cornell university by checking its content to what extent professors participate on it, and compare it with three institutional repositories use DSPACE program which Cornell repository depends on. The study depended on results of interviews of 11 professors in social sciences and humanities to know reasons that led to lake the use of repository of the university. Research reached to: The main reason of professors do not use repository is the weakness of its contents, in addition to professors do not know using methods of repositories, besides weak protection of author's right of scientific works that are in repository.

3 RESEARCH PROBLEM

Basically, this search aimed to identify researching digital repository in economic. The importance of

study insisted on to what extent repository achieved its aims which based on it. the study focused on digital repositories and their importance today, generally and researching digital repository, particularly, and present the previous by analyzing reality of researching digital repositories in economics and what is expected of it.

Today, the technological development with specialized digital repository such in economics one ensure to user benefit with less efforts and less time, so this study seeks to focusing on researching repository in economics in details as follows:

- Identify the digital repositories in economics REPEC.
- Identify the total size of these, saves, magazines, and scientific researches.
- Identify the published articles in patrol.
- Identify the strategy of saving digital sources and managing rights of property.

The foundation stone/bedrock which the study based on is the analytical measure of reality and expected of this repository, today. To assurance the quality of electronic content in these repositories and in economics one, particularly it is necessary to focusing on chances and challenges which surround digital repositories. Therefore questions of this research summarized on following questions:

- What are the infrastructure and the foundation of establishing researching digital repositories in economics?
- What is the size of intellectual production that represents in researching digital repository in economics?
- To what extent it achieved what is expected and is it suitable with its reality?
- What are future planes to develop researching digital repository in economics?

4 STUDY APPROACH AND METHODS OF COLLECTING DATA

To achieve the aims of study researcher used analytical description method because it is optimized with nature of this study which depends on study case method as it depends on study phenomena as in reality and compare it with expected and cares about describing it, in addition to induction of intellectual production/thinking as it is regard as the foundation of HD in all fields and academic social demand which call for to satisfy needs of researcher and academics, also the SOWT analysis was used.

4.1 *study population and its sample*

A sample of study is the sample that presents intellectual thinking/production in Arabic and foreign

languages to recognize directions of research in the field of digital repositories through checking different seven studies includes digital repositories generally and in details, and specialized repositories of different universities, also researching digital repository in economics.

And foundation of establishing of researching repository in economics?

Researcher noticed, when she prepared for researching repository in economics, strong information architecture, also there is caring of saving all works electronically, therefore answers of this question can be summarized in the foundation of researching digital repositories in 3.2-Tools of study.

To achieve the aims of study the researcher presented the Arabic and foreign previous studies of general and special digital repositories and analyzed it and searched on other information that related to it and the internet by using the analysis method SWOT; which used as general strategic analysis tool. This analysis is divided into, as its letters are written in English, S-W-O-T, we can define it as the following:

- 1-strength;
- 2-weakness;
- 3-opportunities;
- 4-threats;
- 5-Results of study.

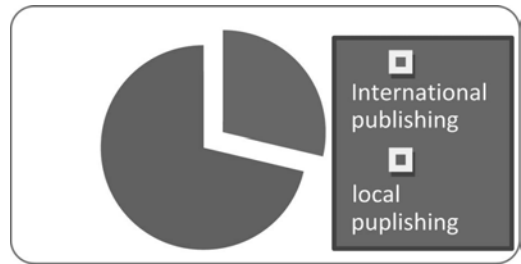
Based on foregoing and analysis of intellectual production of previous studies and comparing between reality and expected of researching repository in economics, we reached to the following results:

4.2 *Results of the first question which included 'what are frustration economics*

The frustration of this repository REPEC is a cooperative effort of hundred volunteers in 87 countries to reinforcement near million searches on science and economics. The center of the project is decentralized and bibliographic data base of papers, articles, books, and contents of program, which volunteers save it, then they use the data that is collected by users service.

4.3 *Results of the first question which included what is the size of intellectual production that's represents in researching digital repositories in economic?*

In RePEc to develop searching and educational process which aim to save intellectual production in bibliographic or textual data bases that is contains 2 million research cuttings such 2300 magazines and 4300 paper, about 4600 authors recorded, and present 75000 of replies on emails weekly. We can describe it as enormous intellectual production. Repository contains high level different articles,



(Shape 1. Size of both international and local publishing in specialized digital repositories).

magazines, books, and searches in economics. Hence the intellectual production's publishing continues to increase in economics in both national and international sides (look shape 1).

4.4 *Results of the third question to what is extent it achieved what is expected and is it suitable with its reality?*

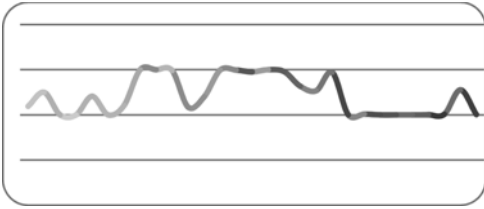
The answer of this question shows purpose and final result which the study built on, it is one of important substrates of this search. All previous assumptions, facts, and studies clarify this repository is useful or not, and it achieved what is expected and this related to reality.

In shape 2, it is noticed that in articles' percent of whole full text there are patrols released their articles in complete text at economy magazine and journal of research business 99.3% except one article does not rise in both because most of patrols are new and have limited production. For example, the economy magazine established at the beginning of 2010 it is published monthly. Also there are six articles which are regarded medium full text (almost half of it published as full text and the other one published as a summary) as it presents 63.5%,55.3%,53.6%,50%,41.2%,40% to them respectively. (Yakel, Elizabeth)

4.5 *Results of the forth question which includes what are future plans to develop digital repository for research in economics?*

Researching repository in economics is considered researching volunteer works in economics and due to supervisors' efforts of digital repository in economics REPEC and what happens on reality today, without doubt, there are enormous efforts, which essentially seeks to develop this repository, as takes in consideration the following:

To answer this question by looking at statistics and studies, we can say briefly the REPEC is enormous repository, it works to develop its scientific



Shape 2. Percent to articles of whole text in order of yield of REPEC.

subjects and forms of benefits constantly. (look shape 2)

- Intellectual production of digital repository in economics are allowed 100%
- Take in consideration save full text to all articles.
- Availability of scientific production that is related to researching in economics without any restrictions on availability of theses, patrols, and conferences.

Intellectual production are allowed inside digital repository for researching in economics and outside universities 10% in line with IP protection laws, fair use laws, and open access.

5 THE SOWT FOUR-WHEEL ANALYSIS OF DIGITAL REPOSITORY IN ECONOMICS

The philosophy of basic principal to get information in REPEC is the index/indicator of content publishers by themselves in REPEC publish descriptive data HTTP or FTP on their website after kheldfold protocol it refer to how to organize metadata archive. The digital repository in economics REPEC has admen who search shown information and provide instructions. According to the previous studies, the results of this research, and the important role of digital repositories this study made detailed analysis that takes the most important points of strength and weakness and the important opportunities and threats which the result of the digital repository in economics and analysis agree/harmonize between reality and expected.

Strengths:

- Open access of all contents of the digital repository REPEC.
- The writer service and personal maintenance which allow to volunteers to contribute on the development of digital repository in economics.
- The biggest academy to discuss economic researches.
- Commission plagiarism to reduce plagiarism the contents of REPEC.
- Addition of Russian language in RPECE and put data base to collective information of social sciences.

<p>Strengths</p> <ul style="list-style-type: none"> Free access The writer service and personal maintenance The biggest academy to discuss researches plagiarism Commission Addition of Russian language 	<p>weakness</p> <ul style="list-style-type: none"> There are no decisions enforcing professors on participation The categories did not participate actually Less experiences in using the digital repository's techniques The lake of articles compared with other works in economics
<p>OPPORTUNITIES</p> <ul style="list-style-type: none"> preserving the account /stock of digital repository REPEC Managing the stock of economic researches well Increase the caring of the digital repository REPEC Sharing sources Develop the means of communication 	<p>THREATS</p> <ul style="list-style-type: none"> The RPECE depends on the support of special projects and volunteers intellectual production /thinking takes long time to be published The RPEC depends on email as a connected mean only

List 1. The SWOT four-wheel analysis of digital repository in economics.

Weakness:

- There are no decisions commit professors on recording their searches and their participations to achieve quality but what has been recorded is volunteer's agrees.
- The categories did not participate actually and did not complete their data.
- There is a category has not experience about information and digital repositories well to participate on huge amount of economic information.
- The lake of articles compared with other works in economics which are not suitable with their history and rate.

Opportunities:

- preserving the stock of digital repository REPEC.
- Managing of this account of researches, theses, and articles efficiently.
- Seeking to increase caring and focusing on this repository and showing its importance. Hence encourage this specialized kind of repositories.
- Participation on sources through digital repository in economics and others repositories.
- Develop and ease the means of communication between volunteers and beneficiaries.

Threats:

- The REPEC depends on the support of special projects and its continue based on contributions of volunteers which reduce the chances of its continue and threats its existence.
- The fact/the reality of REPEC that its intellectual production based on volunteered contributions take long time to revise and publish.

- All economic researching participation, contributions, and general queries about EPCE depend on email which regard the only means of communication.

6 CONCLUSION AND RECOMMENDATIONS

Repositories are regarded one of the types of systems of content management which collect intellectual assets and allow use it to support many/a lot of activities today. We can concluded by clarifying digital repositories generally and digital repository in economy particularly because it present possibilities to save special digital content, efforts of volunteered and searchers in economics, availability of exchange information and experiences in local, national, and universal stander; and contributions in developing all works and researches related to economics and collected it in one digital repository. We reached to important deductions as follows:

- The REPEC represent one of components of receiver system as specialized repository in administration and economics.
- The REPEC covers/present theses, researches, books, magazines, articles and others in administration and economics.
- The REPEC treats since date stage enter and save/store its data and research about it in addition to reports made on.
- The REPEC allows research on theses in bibliographic and full text standards, also it allows research on specific these.
- Research cuttings are 2 million specialized work in economics.
- Theses which are downloaded in REPEC in the first/primary stage 4165 theses.
- There are committees specialized to verify the truth of researching administrated works.

Researcher see that the actual and the previous studies include the research in order to the importance of digital repositories today in all fields generally and digital repositories such the digital repository in economics particularly. She knew, in details, the need of beneficiaries today, therefore she reached to some recommendations which improve the educational process and assurance maximum benefit of information and content; so she suggests the importance of the REPEC achieves some aims as follows:

- It provides strong and different ways of communication which assurance in it the reach of scientific subject/content and support all needs and proprieties.

- Provides support of universities not only special support, it should continue to assurance that people always get on information.
- Informational flow in the digital repository in economics remains usable and available.
- The necessity of keeping up to fast change in technology and better developing of the repository.
- Digitalize it in a shape which supports and meets needs today and in the future.

REFERENCES

1. Albasam, Areeg, Alyami & Huda (2013). the digital repositories for quality assurance of electronic learning's content (LOR). The tree conference of electronic learning and far distance learning in the duration 23–26, Rabi 1st H, corresponding 4–7, February, 2013. Riyadh: sudia.
2. Khaleel, & Hanan Hassan (2012). "establishing repository of learning units to develop skills of creating electronic evaluations and designing banks of questions to the students of the faculty of education—Mansoura university "PhD thesis. Mansoura university. Egypt".
3. Farag, & Hanan (2012). institutional digital repositories and its role to support Arabic content and enrich it on the internet "King Fahad" magazine, book 18, edition 2, page 34–135.
4. Omar, & Iman Fawzy (2011). free digital repositories as one of establishing sources in researching libraries. Analytical study, PhD thesis, faculty of arts Helwan university, libraries and information department.
5. Hendawi Saad "suggested pattern of educational units on the internet within quality standards and its effects on some learning sides to students at faculty of education" PhD thesis at faculty of education Halwan university, Egypt.
6. Digital repository of Mansoura university: study case to digital repository of receiver system of libraries' administration—general administration of libraries, working directory/guide of general administration of libraries, prepared by high studies, searches, cultural relations and libraries. Mansoura university.
7. Savenas Ahmed Mohamed Mahfouz. By: Swan, A. (2006). The Culture of Open access: Researchers views and espouses. In N. Jacobs (Ed.) Open access: Key strategic, technical, and economic aspects (pp. 6572-). Oxford: chandos.
8. Abdu El-megd bu Azaa the directions of Arabic researchers to open archives and free courses on the internet: Arabic professors at sultan Qaboos university, pattern cybrarians journal (September, 2006) on: http://journal.cybrarians.info/index.php?option=com_content&view=article&id=528:2011-08-22-03-13-22&catid=120:2009-05-19-11-31-27&Itemid=74.
9. Brown, H. & Abbas, J. (2010). "Institutional Digital Repositories for Science and Technology A View from the Laboratory. Journal of Library Administration. 3, pp. 81215.
10. Connolly, P.M. Institutional repositories: Evaluation the reasons for non-use of cornell University's installation of Dspace. D-Lib Magazine, vol. 13, n.3/4, (March/April 2007). Available in: <http://www.dlib.org/dlib/march07/davis/03davis.html>.

11. Lynch, Clifford A. Academic Institutional Repositories.—D-Lib Magazine, September 2005, Volume 11 N.
12. From: <http://www.dlib.org/dlib/september05/westrienen/09westrienen.html>
13. Yakel, Elizabeth ... et al. Institutional Repositories and the Institutional Repository: College and University Archives and Special Collections in an Era of Change.—American Archivist. Vol. 71, N. 2(Fall/Winter 2008). 323–349. URL: <http://www.metapress.com/content/c7t344q22u736lr2/>.
14. Harnad, S. & Mc Govern, N. Institutional Repositories Success Is dependent upon mandates. Bulletin of the American Society for Information Science and Technology, V. 35, Issue 4, pages 27–31, (April/May 2009). Available with: <http://onlinelibrary.wiley.com/doi/10.1002/bult.2009.1720350410/abstract>.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Social network applications: Critical review study

Majed Mohammed Abu Sharha

Department of Information Science, King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT: Social media networks is one of the most important techniques that has existing influences on our communities and most widely spread in the process of communicating with others as well as speeding of news dissemination, exchange of information, transfer of knowledge and experience. Recently a lot of organizations adopted social media networks in establishing, forming and sharing of knowledge among its members.

In this paper we will review the definition of social media network types, and their most important positives sides and characteristics, some of their drawbacks and highlight the roles played by these networks in education, marketing and e-government, as well as you will learn about the most important aspects of privacy in social networks. Finally, the paper will provide a set of results and statistics.

Keywords: Social network, Social media, Web 2.0, Facebook, Twitter

1 SOCIAL MEDIA NETWORKS

Web 2.0 is considered as the most important application of social media networks and its importance comes from allowing interaction among users, allowing communication among members and providing services in ways that is unprecedented.

Facebook is the most famous of these networks that offers a service to join one of the available groups which are classified according to their topics and to make friends as well as to get a new knowledge and share resources among members. Another site is Twitter, which provides micro-blogging service that allows users to send Tweets updates on their status within of 140 characters per message as maximum. Further there are a lot of other social networks like MySpace, whose users are estimated with more than one hundred million. MySpace provides interactive visual and instant messaging services among registered users in addition to other services like blogging and sharing photos and video files. The Flickr is also a site that allows sharing photos and video between users, with possibility to save and organize.

Social media networks are defined as: “Tools that support social relations between individuals through the Internet by using many means such as forums of dialogue, files exchanging, conversations, blogs, and summaries of sites in the virtual world. All these tools are working together to enable individuals to control their time, activities, and their relationships with each other” (Andurson, 2005).

2 SOCIAL MEDIA NETWORKS TYPES

1. Blogs: World Press Blogger, Twitter, TypePad, Live journal and Snapchat.
2. Social Networks: Facebook, Myspace and Linkdin.
3. Events: Tweetvite, Meetup and Eventful.
4. Social News: NowPuplic, Reddit, Technorati and Newsvine.
5. Photo Sharing: Instgram, Flickr, Picasa and Phtobucket.
6. Video Sharing & Streaming: YouTube, Lives-tream, Sevenload and Metacafe.

3 SOCIAL MEDIA NETWORKS ADVANTAGES

1. Improving of employment opportunities for young people.
2. Reducing of barriers to communication.
3. Increasing of concentration on technology.
4. A new source of urgent and immediate news.
5. Awareness enhancing with social, cultural and political values.
6. Increasing of self-learning opportunities.
7. New marketing channels (Fantookh, 2015).

4 SOCIAL MEDIA NETWORKS DISADVANTAGES

1. Ease of unethical practices.
2. The rapid spread of rumors and defamation.

3. The indifference and irresponsibility of some users.
4. Concern for other work rather than study, business or life tasks.
5. Arising of frequent intellectual conflicts and debilitating that made individuals exhaustive.
6. Violation of privacy, extortion and fraud.
7. Impacts on children and adolescents.
8. Prone to social isolation and weak family bonding.
9. Opens doors to opinions of non-professionals.
10. Frequent accidents and traffic violations because of drivers' preoccupation during follow-up of these media. (Fantookh, 2015)

5 MODERN SOCIAL MEDIA NETWORKS MEANS CHARACTERISTICS

Almahmoud, 2015 mentioned in his doctoral dissertation on criminal responsibility for the misuse of social media networks, that there are multiple properties of social media networks especially the modern ones, which are as follows:

1. Absolute freedom from restrictions.
2. Interactivity and after the reaction.
3. Fragmentation of audiences (Media Fragmentation).
4. Multiple forms of media publishing.
5. Absence of synchronization which means the lack of need for sender and receiver to be present at the same time, receiver can get content at any time he wants.
6. Proliferation and universal access.
7. Difficulty of departure.

6 SOCIAL MEDIA NETWORKS ROLE

Social networks play many roles in most aspects of life: economic, political, education, medicine etc, and its impacts top the ongoing events in the world because most of individuals are linked to those networks. We are going to review some of the main areas where social networks are used:

7 SOCIAL NETWORKS ROLE IN EDUCATION

The use of social networks in education help a lot in developing educational process, it has led to make a positive impact on teachers' performance of and learner method.

Alhazani, 2013 mentioned that social networks play an important role in providing students with information that serves most of disciplines, the

following are the most important services offered in educational field:

1. Exchanging of e-mails.
2. Social networks allow files transferring that include texts, software, pictures and sounds among students.
3. Providing of variety and updated information in a best way when compared to other means of communication.
4. Diversity of services provided by social networks drive a spirit of enthusiasm and motivation among learners.
5. Social networks promote a powerful development of students' scientific creativity.
6. Social networks provide an easy mechanism for students and teachers to publish their work.

But there are still many obstacles to use of social networks in education, including:

1. Reluctance of some teachers towards using of this technology.
2. Slow change rate in bureaucratic systems.
3. Frequent change, lack of stability and fixing of sites and links that link between different sites on social networks.
4. Learner needs a lot of time in some social network sites in order to get pictures and sounds.

8 SOCIAL NETWORKS ROLE IN APPLICATION OF E-GOVERNMENT CONCEPT

Many governments tend to change the prevailing thought in the past that citizen goes to government but with the era of social networks, there emerged a need to interact with citizens and open channels of communication with him.

According to the governmental policy draft issued by Information Technology Authority of Oman (2013) to engage electronically and to use of social networks in governmental sector through necessary work on developing regulatory public policies for using of social networks within the e-government websites. These policies will enhance partnership and cooperation between these sites and their beneficiaries, in addition to increasing the effective participation of citizens and their sense of their role in participating within the e-government programs.

9 THE IMPORTANCE OF USING SOCIAL NETWORKS IN E-GOVERNMENT

The social networks that used by the government websites are considered as the most important

means of communication at this times as it offers many advantages that support the process of communication between the government institutions and citizens, including:

- Access to citizens is increased, and developing of government communication process.
- It is considered as one of the favorite communication tools by many citizens.
- Meet the citizens' expectations in regard of modern services and enhancing of its reputation.
- Enhancing transparency.
- Strengthening relations with citizens, partners and stakeholders.
- Strengthening government's performance by focusing on the process of communication and its improving.
- Rapid access to responses of citizens and their contributions.
- Ability to access to a certain class of citizens on specific issues.
- Reducing dependence on applicable media means and facing inaccurate press coverage.
- Creation and development of electronic contents written by the citizens.

10 SOCIAL NETWORKS ROLE IN MARKETING

Marketing is the most important field that benefited greatly from rapid developments in social network applications, which represent the right environment for the provision of marketing services sought by many companies. Social media marketing is the activity to exploit social networks by corporate in marketing purposes through identifying and analyzing conversations and participations and to initiate social interactions within communities in order to use.

We also mean by marketing through Social media networks and sites the using of those tools and sites of social media, such as: Facebook, Twitter, Instagram, Snapchat, Youtube, linkedin etc, for marketing of services, goods, or any type of commercial activities in fields of business in order to achieve marketing business objectives which determined carefully.

11 PRIVACY IN SOCIAL NETWORKS

Privacy is an important factor in social networks work as we have seen that most of those sites and applications allow using of personal data and images by users from which appears the importance of privacy. Question that always

comes to mind, how users of social networks can participate and benefit from services provided with control of privacy as well? This is varies depending on purpose of participation in social networks, as some may take part for purpose of communicating with others or in educational environment, while others may be with marketing purpose. According to the purpose of privacy control method varies.

Prominent security risks in social networks are as follows:

- Phishing: obtaining of private users information, whether personal or financial information via e-mails or Web sites that appear to come from reliable companies or governmental financial institutions.
- Identity theft: the attacker spoof and falsified user's identity and pretend as if he was a person or user.
- Spam: undesirable mails, as are so many users of social network are susceptible to problem of dumping.
- Stealing of information and modifying it: are mostly done by third party.

Careful shall be due to educate all users, whether individuals or institutions to maintain their privacy and increase security awareness.

12 RESULTS AND STATISTICS

Digital world has witnessed great development in the past few years, the thing that contributes to an increase in using of social networks sites until the number of users jumped in 2014 to 2.5 billion person around the world.

- As per statistics published by "skynews arabia" site, site Facebook ranked as forefront in terms of its global use, where the number of active users on a monthly basis surpassed 1.6 billion user.

13 CONCLUSION

We conclude from the foregoing that the social network sites caused a direct impact that become clear and concrete in several fields and behavior and attitudes of individuals as well. But the largest importance must be dedicated to optimal use of social network applications and using them appropriately in those areas utilizing of social networks' advantages and potential and effective and influential capabilities. We must deal with it as a tool for communication and dissemination of

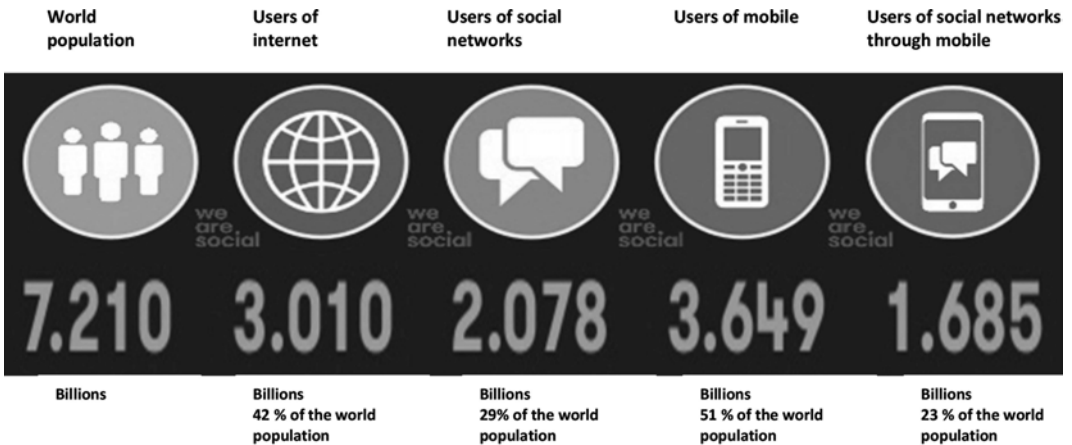


Figure 1. Global statistical indicators of social networks using (courtesy; we are social).

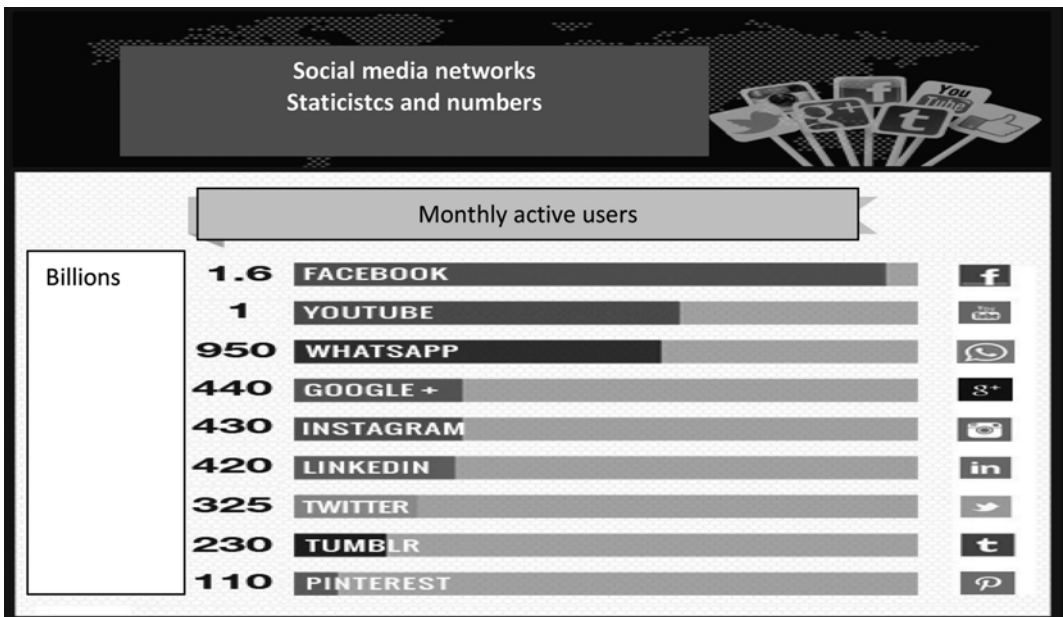


Figure 2. Monthly active users (source: “skynews arabia” site).

the best values within our communities, without forgetting their real role. We look forward that those networks must be one of the transition tools of knowledge community. Before creating a page or account on social network sites, whether personal or institutional, we have to develop and plan a systematic, integrated and comprehensive strategy

based on a deliberate plan that aims to highlight important issues we want to discuss. Therefore we work to resolve them through participation of specialized professional groups or directly communicate them to officials and decision-makers to resolve those issues, and that’s what we started to see lately.

REFERENCES

- Arici, Jibril, al-Dossari, Selma (2015). Social networks and values: Analytical vision.—Amman: Aldar Almanhija.
- Alfantookh, Abdul Qadir (2015). Social Networks: Impact and future.—social networks and intellectual security conference. Riyadh: Saudi Computers Society, 02/11/2015.
- Almahmoud, Mohammad (2015). criminal responsibility of modern social media misuse of Riyadh: Dar Jawasr for heritage and publishing.
- Alhazani, Nora (2013). The effectiveness of electronic social networks in developing of teaching and learning process among students of the Faculty of Education at King Saud University.—International Journal of Educational Research/University of the United Arab Emirates Issue 33.
- Anderson, T. (2005). Distance Learning. Social Software Killer ap.-available at: http://auspace.athabascau.ca:8080/bitstream/2149/2328/1/distance_learning.pdf.
- Ellison, N.B. & Boyd, D. (2007). Social network site: Definition, history, and scholarship. Journal of Computer-Mediated Communication. -available: <http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00393.x/epdf>. www.skynewsarabia.com, www.wearesocial.com.
- Information Technology and Communications Authority, Sultanate of Oman, (2013). Government policies draft for electronic participation and use of social networks in governmental sector issued by Information Technology and Communications Authority, Sultanate of Oman.
- Telecommunications Regulatory Authority (2011). Guidelines for using of social network tools in United Arab Emirates government websites.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

The effective use of social networks to support knowledge acquisition process at KAU, Jeddah, Saudi Arabia

Maysa Ibrahim Yousef & Ebtesam Hussain ALZahrani

Department of Information Science, King Abdul-Aziz University, Jeddah, Saudi Arabia

ABSTRACT: This article aimed to assess the effectiveness of the use of social networks to support knowledge acquisition processes in Saudi universities' education by specifying the programs that are frequently used by graduate students to get new information and knowledge and reveal the relationship between the use of social networks and its impact on the educational achievement of graduate students by adopting a descriptive method. It has been explained through two questionnaires:

- Demographic variables which in fact are academic achievement.
- Questionnaire includes a list consisting of 29 programs or technical activities that are employed in education and teaching.

64 students were chosen randomly to apply these programs on them. The results showed that Facebook, Google, e-mail and YouTube were much used programs to acquire latest information and knowledge while programs related to (RSS), podcasts, electronic curriculum and media sharing emerged as the least used. However, it was understood that there is a significant correlation statistically between the use of social media and academic achievement as the value of the correlation coefficient reached (0.55); a statistically significant value at the significant level (0.01).

Keywords: Digital Social Networks, Academic Achievement, Knowledge Acquisition, Knowledge Abstraction

1 INTRODUCTION

The digital social networking has been playing an active role in providing a lot of information, attitudes and trends, as well as these networks are distinguished with several features such as participatory, interactive and intangible permanent presence. The increasing number of participants in these digital networks, particularly the Arab youth has led to an escalation of its impact and role in the student community and their levels of academic achievement. The digital social media has been able to create new knowledge, freedom of expression and exchange of ideas among university students.

The social network approach holds that the behavior of an individual is affected by the kinds of relations, or technical ties, and networks more than by the norms and attributes that an individual possesses. The social, informational, or material resources that two individuals exchange characterize their ties. In social network analysis, these resource exchanges are termed "relations." Some positive and negative relations are assumed to be related to an individual's performance. Researchers

empirically demonstrated that friendship and advice relations were positively related to a student's academic performance and an employee's job performance. On the other hand, the effects of an adversarial network were negatively related to performance [7, 8]. It seems worthwhile to investigate the effects of the three social networks on student performance online and offline.

Since "Randy Conradz" built the first pillar of social networking sites by developing the first site to communicate with his friends and classmates which he named "classmates.com", the social networking sites have become popular among the university students as it succeeded in presenting itself as an interactive, unique and useful media at different levels of educational work on the web since its inception, and in the light of what we mentioned above there emerged a desire to know the effectiveness of using social networks to support the process of acquisition of knowledge in King Abdul Aziz University.

The social network formed by these students was different from that of distance learners since the latter developed their relationships mainly through online interactions. Actually, there were three

sessions during which the students could develop their networks—in the AMIS class, before and after the class, and in the forum. Since the class period was only three hours per week, we might conjecture that most of the friendship and adversarial networks developed after the class. In the AMIS class, most of the discussions were one (lecturer) to many (students). Therefore, although students were motivated to show their knowledge during the class, the advice network could not develop. However, on the forum, the discussions were many to many. Everyone was free to express an opinion and knew the teacher was watching to see how valuable were the opinions or information they provided to all the members of the forum. The advice network could naturally evolve over time. This might explain why the advice network centrality is the best determinant for explaining performance variance.

2 RELATED WORKS

Becoming aware of the literature of intellectual production, we reached to some studies relevant to our current curriculum that is effective use of social networks in supporting the process of acquisition of knowledge in the education of the King Abdul Aziz University in Jeddah. In the work entitled: scientific communication within the social networking environment in 2014 [Wardah Museibih, 2014]; the work aims to disclose of the nature of the relationship between social networks and scientific communication, knowledge transfer and its flow, and to be aware of the role of these networks to strengthen scientific communication between professors, researchers and students at universities. The main findings illustrate the impact of social networks on the patterns of scientific communication in academic environments continues to grow, not only in terms of quantity of knowledge that is available on the social networks, but the quality of advanced services offered by the social networks to students and researchers. With the transformation of the container paper technology to the electronic container individual and collective practices for undergraduates and associated with the creation and transfer of knowledge have led to a change in production methods, processing and organization. The revolution that we are experiencing now is related to new concepts and tools such as: multi-media, social networks, hypertext, and knowledge structure.

With respect to the second related work entitled: Social Networking as an Alternative Environment for Education 2012 [Andrei Stanciu, others, 2012], the authors aimed to provide a typical proposal for the implementation of the use of Facebook in the process of education and knowledge transfer in higher education. It also aims at knowing the

reason why university students join social network sites. The main conclusion of this work proved that social networking sites have become very popular among university students and regarded as a tool for education and knowledge transfer, as well as the flexibility of access provided by the new technology, but it has large effect on the frequency of access to social networking sites anywhere and at any time.

On other hand, in the work entitled: University Campus Social Network System for Knowledge Sharing 2012 [Zhao Du, others, 2012], the authors try to design and implement a social networking system within university campus with a focus on knowledge-sharing mechanism in this modern system, and to ensure that this social network is effective in spreading the sources of digital information through this knowledge sharing process. The major results of this study indicate that the personal social networks provide the users with the basic channel for exchange of knowledge through the social networking system within the university campus, and significantly contributes to the construction of another channel for the exchange of knowledge for students within campus. The system allows controlling access to digital resources by students for the exchange of knowledge through it. It also provides a multi-scale assessment process for digital resources, creating more than one way to find out the most useful and effective quality of digital resources with minimal effort.

In [Patient Rambe, 2011] entitled: Networking Sites on Academic Relations in the University 2011, the authors aim to detect the effect of the use of social networks on the acquisition of knowledge and academic relations between lecturers and students of the first year courses in Information Systems (IS) in South Africa's medium-sized universities, with a focus on those relationships in the Facebook environment. The main findings of the study were: for some academics, the social networks are educational and cognitive tools that provide students with assistance when they face difficulties in learning; it also helps in carrying out social monitoring. Some students at the university are using Facebook to express their grievances and dissatisfaction with some of the management practices in their colleges. However, educators and faculty members use social networking for information technology, as well as educational and cognitive purposes.

Finally in [Christian R. Østergaard, 2007] entitled: Knowledge Flows through Social Networks in a Cluster: Interfirm versus University-Industry Contacts 2007, the researchers aims to study the effectiveness and usefulness of social networking sites as channels for sources of knowledge among industrial entities and research scholars at local universities, in addition to reveal the relationship

between the acquisition of knowledge and different characteristics of engineers in the factories as well as research scholars in universities. Then the study compares and analyzes the extent of dependence of universities and industrial entities on social networks. The main achievements of this work show the changing jobs and places, switching of an employee between companies or of a student between universities is a kind of knowledge flow. In addition to that, changing of jobs contributes automatically in creating an informal channel for the flow of knowledge in the present and the future. The previous joint work experience may become more important in creation of social networking used for the acquisition and flow of knowledge. As the working experience on the same project together contributes a lot in the transfer of knowledge. Many engineers make unofficial contacts with research scholars in universities to acquire some knowledge from them. The social networking spreads knowledge among companies with each other on the one side and between universities and companies on the other. It is also possible to achieve the process of spreading and flow of knowledge between the alumni through formal cooperation as well as through informal cooperation through informal social networks.

3 PROBLEM STATEMENT

The problem of the study lies in finding out the scope of effective use of social networks in supporting knowledge acquisition process at King Abdul Aziz University in Jeddah. So, the main queries of this article lie in:

- What programs popularly being used by graduate students for acquiring new knowledge and information? To answer this question, we made a calculation of averages and standard deviations to know the extent of use of each program contained in a list of 29 programs or technical activities that are employed in education and teaching.
- Is there any relationship between the use of social media and its impact on the educational achievement of graduate students? To answer this question and to reveal the statistical significance between the two variables, we used Pearson correlation.

4 METHODOLOGY OF SOLUTION AND DATA COLLECTION

Method of the study is descriptive and correlative as it is the most appropriate research method.

Every one of Abidat, Adas and Abdul Haque (2016, Page 296) mentioned that this method means studying the phenomenon and to identify the factors affecting it. This method is not limited to just a description of the phenomenon but it also interprets and analyze the phenomenon to characterize the reality of the role of the use of social media in creation of knowledge and exploring its impact on academic achievements among graduate students.

Thus, the population of study consisted of all graduate students of scientific and humanitarian disciplines at King Abdul Aziz University, and the total number of them reached 400 students. The study sample consisted of graduate students who were selected randomly and their number could not exceed 64 students in view of the difficulty of access of researchers to all the students due to the different academic circumstances and schedule of lectures.

4.1 *Tools of the study*

To achieve the objectives of the study, we used questionnaire as it suits to the objectives of the study. The questionnaire consisted of two main points. The first point is demographic variables, which are in fact academic achievement and the second point included a list of 29 programs or technical activities, which are in use of teaching and education.

4.2 *Limits of the study*

Objective limits: the study was limited to the following objective limits; the effective use of social networks in supporting process of acquisition of knowledge in the education system of King Abdul Aziz University in Jeddah, Saudi Arabia.

4.3 *Spatial limits*

The spatial boundary was limited to graduate students of King Abdul Aziz University in Jeddah, Saudi Arabia.

5 MAJOR FINDINGS OF THE WORK

With respect to the main findings of the first question and its analysis: What are the programs popular among graduate students to acquire latest knowledge and information?

The authors found that the percentage of the use of programs popular among graduate students to acquire new knowledge from their perspective was averaging as its average reached (3.38) with a standard deviation of (1.02).

It is evident from Table 1 that the most frequently used programs were E-mail, Facebook, YouTube, Google+ Snapchat, Google Calendar and Google

Table 1. Average and standard deviation of the grade of programs usage to acquire knowledge. The grade is in ascending order (ascending = 64).

No	Programs	Grade	Average	Standard Deviation	Level of use
57	Use of Email	1	4.19	0.88	High
63	Facebook	2	4.18	1.02	High
65	YouTube	3	4.15	1.08	High
48	Goggle+	4	3.90	1.13	High
47	Snapchat	5	3.88	1.19	High
49	Google Calendar	6	3.82	1.22	High
50	Yahoo Calendar	7	3.75	1.20	High
51	Google Calendar	8	3.72	1.32	High
60	Images and Flashes	9	3.65	1.09	High
62	WhatsApp	10	3.61	1.12	High
66	Wikis	11	3.60	1.15	High
67	Blogs	12	3.56	1.03	High
64	Twitter	13	3.55	1.08	High
58	Internet Conversation	14	3.52	0.89	High
46	Instagram	15	3.38	0.91	Medium
45	Word press	16	3.35	0.96	Medium
52	Flicker	17	3.33	1.20	Medium
54	Daily motion	18	3.39	1.20	Medium
55	Slide share	19	3.25	1.44	Medium
56	Scribed	20	3.13	1.35	Medium
68	Social favorite	21	2.89	1.36	Medium
69	Virtual video	22	2.88	1.22	Medium
70	Media sharing	23	2.85	1.25	Medium
71	Podcast	24	2.83	1.36	Medium
72	RSS	25	2.80	1.40	Medium
73	Dialogue Group	26	2.86	1.36	Medium
59	Interactive	27	2.82	1.25	Medium
61	Interactive video and fixed movie	28	2.69	1.20	Medium
45	MySpace	29	2.65	1.26	Medium
Grade of the use of the programs			3.38	1.02	Medium

Picasa. Perhaps these programs were most widely used by public being the means of social communication between students in social matters and to exchange of dialogue. Arithmetic averages ranged between 4.19–3.52. This is in addition to previous programs, Yahoo Calendar, Google Calendar, images and flashes programs, WhatsApp, Wikis, Blogs, Twitter, Conversation Internet Relay Chat IRC and Instagram.

Perhaps the reason is that the practical life activities focuses on use of such ready-made programs in daily social process and take advantage of them in academic and educational matters, as well as some of the programs have become familiar topics for most students to depend on them.

The other programs and applications emerged with medium average use ranging from 3.38 to 3.65. This result may be due to the different skills owned by the students as well as due to availability of other programs that serve the purpose better. And the programs that were used at a low average included the use of Dialogue Group/Discussion Group, Media Sharing, Twitter, RSS (feed)

websites summaries intensive, Podcast audio and visual broadcasting, Virtual video conferencing, and use of Social Favorite. These programs were used at a lower average because of their modernity and difficulty of use in the study and teaching curriculum at graduate level in different disciplines in terms of subjects and skills that are in focus as well as these programs need specialized knowledge of their use in educational aspects.

The achieved results of our study agreed with the findings of the study that was carried by Andrei Stanciu & Others [Social Networking as an Alternative Environment For Education, 2012], which revealed that commonly used softwares in social networks is Facebook, which is used for the dissemination of knowledge. On other side, the main findings of this study differed from the findings of Andrei Stanciu & Others [Social Networking as an Alternative Environment For Education, 2012] study that revealed Twitter is not a well-known program while the current study found that Twitter is being used more highly.

Table 2. Results of Pearson correlation between the grades in use of social networks and cumulative rate among graduate students.

	Percentage of use of methods of social networks
Cumulative rate (achievement)	**0.55
Number	64

With respect to the results of the second question and its analysis: If there is any relationship between the use of social networks and its impact on educational achievement among graduate students.

The researchers found that there was statistically significant relationship between the grades of students in cumulative and variable of the use of methods of social networks at the level of ($\alpha = 0.01$), as higher the methods of social networks to acquire and create the knowledge the greater the educational achievement representing a cumulative average. The reason for this, perhaps attributed to that grades obtained by the student in higher studies becomes influenced by the work carried out during the semester which is known as work of the year represented by the costs, duties, reports and scientific papers. And certainly, being excellence in work of the year and submitting assignments' solutions with high professionalism depends on how much the student refers to sources and internet. The more the student uses technology in his home works the better result he will get. Finally, we assure that the present study is alone in this variant, researchers did not find a study agreed or disagreed with the results of the study of the differences attributable to the rates and educational achievement.

6 CONCLUDED REMARKS AND MAIN RECOMMENDATIONS

We conclude from aforementioned research that the average of use of the programs that are common with graduate students to acquire knowledge is medium, and its arithmetic average reached 3.38 with a standard deviation of 1.02 And that the most commonly used programs were Email, Facebook, YouTube, Google+, Snapchat, Google Calendar and Picasa. The average of their use ranged from 4.19–3.52 with standard deviations of 0.88–1.32. And the programs that were used at a low average included the use of Dialogue Group/ Discussion Group, Media Sharing, Twitter, RSS (feed) websites summaries intensive, Podcast audio and visual broadcasting, Virtual video conferencing, and use of Social Favorite.

The study also found that there was statistically significant relationship between the grades of students in cumulative and variable of the use of methods of social networks at the level of ($\alpha = 0.01$), in the sense that the greater use of the methods of social networks to get the knowledge the greater the educational achievement representing a cumulative average.

Accordingly, we recommend the need to work on the use of applications of second-generation technology in teaching curriculum in higher studies; we emphasize the importance of creating mailing groups for educational purposes; and employ social networks such as Facebook in blogs related to study; and create links to other pages while teaching topics of the curriculum; and upload files or place it on the Internet during the study as the results revealed that the average of use of these programs are generally medium. Also, we recommend working on the training of faculty members and developing skills relating to use of applications for dissemination of knowledge such as blogs or wikis, and evaluate the posts of students based on clear criteria on for blog or wiki, and the ability to bring and browse students' posts in blogs via RSS.

REFERENCES

- [1] Adas, Abdurrahman, Abidat, Zuqan, Abdul Haq, kaid. "Scientific research, its concept, tools and methods". Available at: <http://www.daralfiker.com/node/6497>
- [2] Wardah Museibih—"Scientific communication within environment of social networks" Cybrarians Journal, Issue: 36, December 2014. Available at: http://www.journal.cybrarians.org/index.php?option=com_content&view=article&id=675:socialmedia&catid=270:studies&Itemid=99 (1 Dec, 2015).
- [3] Andrei Stanciu, others.—"Social Networking As An Alternative Environment For Education", 2012. Available at: ftp://ftp.repec.org/opt/ReDIF/RePEc/ami/articles/11_1_4.pdf (1 Dec, 2015).
- [4] Zhao Du, others.—"University Campus Social Network System for Knowledge Sharing", 2012.—Available at: <http://www.doiserbia.nb.rs/img/doi/1820-0214/2012/1820-02141200055D.pdf> (1 Dec, 2015).
- [5] Patient Rambe.—"Networking Sites on Academic Relations in the University.—Journal of Information Technology Education".—Vol 10, 2011.—Available at: <http://www.jite.org/documents/Vol10/JITEv10p271-293Rambe981.pdf> (1 Dec, 2015).
- [6] Christian R. Østergaard.—"Knowledge Flows through Social Networks in a Cluster: Interfirm versus University-Industry Contacts", 2007.—Available at: <http://www3.druid.dk/wp/20070019.pdf> (1 Dec, 2015).
- [7] Baldwin, T.T., Bedell, M.D. and Johnson, J.L. *The social fabric of a team-based M.B.A. program: network effects on student satisfaction and performance. Academy of Management Journal*, 40(6): 1369–1397, 1997.
- [8] Sparrowe, R.T., Liden, R.C. and Kraimer, M.L. *Social networks and the performance of individuals and groups. Academy of Management Journal*, 44(2): 316–325, 2001.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

The role of open access in supporting the digital repositories activities

Maysa Ibrahim Yousef

Department of Information Science, King Abdul-Aziz University, Jeddah, Saudi Arabia

ABSTRACT: This study aimed to determine the role of open access to information in supporting the digital repositories events where the digital repositories previously acquired great importance especially those between universities and research centers. Studies have confirmed the importance of repositories in forming the regional clusters between universities and libraries so the direction of open access of digital information has emerged in the past few years as a main development in the world of scientific communication which returns on the entire community from facilitating the flow of the information freely without legal restrictions. This study has observed some of the obstacles that prevent the utilization of digital repositories and prevent open access. It has proved the weak presence of the policies which described the placement processes to enable open access for different materials by 57% from the total Arab repositories. It was found that the most available items of placement in the Arabic repositories is materials which may not be used fully as the placement policy is optional in international and Arabic repositories by 66.5% and mandatory by 33.3% from the total international repositories.

Keywords: open access to information, SWOT, RRS technology, Digital repositories, Digital information, technical restrictions, scientific communication

1 INTRODUCTION

Scientific and technical information has great importance in information society it is the primary and main material for achieving development in different fields of life because of its great benefits which help in taking the right decision as it witnessed great development and wide growth on the internet, which is considered as one of the most successful means to transfer and exchange them, for the speed of communication and the storage capacity with the possibility of accessing these information in different forms and types easily.

Scientists who developed this technology “reconsidered the traditional systems of scientific publishing and create modern communication systems which enable providing the information and knowledge freely without legal and technical restrictions, and this is the basis of open access philosophy, the most significant aspects and manifestations of this movement, the digital repositories, which are considered a new model of free scientific communication which allows the availability of intellectual production in its full content and free of charge. This movement based on the principle of re-owing the different scientific sources to the researchers through electronic publishing via the internet without intervention from commercial publishers, this is a useful tool that seeks to publish scientific outputs and electronic sources

to improve the opportunities of free access to different sources, increasing their presentation to researchers as they are considered as digital repositories which contains a digital balance of scientific publications exchanged in different circles that are indexed with metadata technology which returns on the entire society with facilitating the flow of information.

2 PREVIOUS STUDIES

Becoming aware of the literature of intellectual production, we reached to some studies relevant to our current curriculum that is: The role of open access in supporting the digital repositories activities. The first study of *Ballou Amna* entitled: open institutional archive and open access to scientific and technical information, 2014. This study aims to identify the movement of open access to information and monitoring the different ways of access, also aims to identify the open archive and its importance for researchers, international community and its role in saving the digital sources and accessing them freely without financial or technical restrictions. The major findings of the study proved that free access to information in open archive increases the knowledge rates of intellectual works of researchers and enables the researchers to provide the research drafts through

repositories to prove priority and to get comments from knowledgeable colleagues and settling the review of researches [1].

The second study authored by *Ezzedine, Nagda, Zineb* entitled: University libraries and initiatives to achieve free access to information and exchanging them through the electronic environment between the efforts of achievement and its obstacles, 2014. This study aims to highlight the extent of university libraries awareness about the importance of free and open access to information in academic research environment and identify the most significant contribution and initiative in support of achieving the principle of free access information without financial or legal restriction. The main achieved results of this study clarifies that the libraries have adopted the system of free access information as a competitive and alternative system to the traditional system which based on linking access to information with legal and financial restrictions. Free access to information has a significant impact on library services and their economies. It is observed the great efforts and strong will to spread the movement of free access information between the researchers, students and university professor [4].

The third study of *Ahmed Ezzat* entitled: Freedom of exchanging information (legal study), 2013. This study aims to clarify the rights of individuals in free access to information and exchanging them as it considered one of the most important mechanisms to promote and support the practice of other rights of different types. The right of knowledge is the other side of freedom of expression, also aims to provide the suitable means for the flow of information, views, and ideas between state institutions, individuals and different civil society institutions. The major findings of this study illustrate that the free access to information helps in the promotion and protection of economic, social and cultural rights, pursuant to the principle (rights are indivisible) and that the right to access the knowledge and the freedom of exchanging them is not only a right in itself but it is a tool to activate the practice of other rights, and respecting the freedom to access information is necessary for scientific research and creative activity, as it cannot enjoy the freedom of scientific research and innovations without the information being available as they are the basis for every research and creative activity [5].

The fourth Study of *Chimezie Patrick Uzuegbu et al* entitled: Digital Librarians and the Challenges of Open Access to Knowledge: The Michael Okpara University of Agriculture (MOUAU) Library Experience—2012. The objective of this study aims to apply the initiative of free access information in libraries and research centers in Nigeria such as AGORA program of United Nations, which enable the developing countries to benefit from a

wide range of researches, books and scientific articles, facilitate access especially in the fields of food, agriculture, environmental and social sciences. The achieved results indicate that free access to information through AGORA program helps many countries which joined the same program such as Kenya became a tool that beneficiaries depend on it for development such as the countries which failed to join AGORA program like Nigeria. This country Attributed non-joining to lack of knowledge and awareness about the importance of free access to information, the weak technological infrastructure, inadequate funding for scientific researches and the poor state of information technology facilities [8].

In the fifth study of *Fatma Mahmood El Noor* entitled: open access to information in university libraries (case study of the digital repository of the faculty of science in the University of Khartoum, 2012); the author aims to analyze the advantage of the role of free access to information and discovering the most important problems facing students, also aims to find out the type of information received by the beneficiary and its relationship with free access and the extent of contributions of free access information, their types and developed forms in university libraries. The major conclusion of this work reached to an important fact that assures free access to information has a significant role in publishing science and its development in all of its branches, also the study demonstrated that libraries have a great role in supporting free access processes. From the results of this study, the lack of Arabic contributions for setting approved standards contrary to the advanced countries which making clear efforts, also Arabic libraries significantly failed to participate in the development of free access process and supporting it [7].

With regards to the sixth study authored by *Ahmed Ebada* entitled: digital repositories for academic institutions and their role in educational and research process (2011). This study aims to examine more than 50 digital repositories according to the order of the cyber metrics lab, and this for the methods of research, retrieval and access to information from different sources. The researcher concludes that there are many repositories such as repository of king Fahd university of petroleum and minerals, repository of Queensland university of technology, also repository of CERN document server, that seeks to provide free access to information after it has set some restrictions on the access of different materials, as process of free access to information are limited on its participants only. The study confirmed that digital repositories form an essential part in the regional clusters between libraries and universities as it provides the beneficiaries with the possibility of using the sources and taking advantage of them [3].

Finally, with regards to the last study that is searched by *Kylie Pappalardo and Anne Fitzgerald* entitled: *A Guide to Developing Open Access through Your Digital Repository*. 2007. This study aims to clarify many of the principles that can be followed to achieve free access to information through digital repository, which guarantee that the digital repository or the institution is in a strong and advanced position between the other digital repositories. This work reaches to achievement of free access to information inside digital repositories in a good manner will help in attracting new members and students, also helps in maintaining the rare intellectual resources and assets without blocking the access to it. Also the author recommends setting a guide to help different research and educational institutions, and contains many principles that support and assist in the implementation of free access processes to information inside digital repositories [9].

3 PROBLEM STATEMENT

Previous studies have discussed the importance of digital repositories in spreading awareness among societies and the delay of Arabic repositories compared with global counterparts, which persist in providing free and complete access to its different materials, and the role of free access information in the progress of nations, here we can say that the problem of this study is to find out the role of free access in supporting the digital repositories events?

4 RESEARCH METHODOLOGY

This work used SWOT analysis, which is an analytical method to find out the strength and weakness points as well as the opportunities and threats based on analyzing the internal and external situation for any organization through the following four items: strength points, weakness points, opportunities and threats. Whenever the analysis is more accurate, the organization will be in better position to exist for a long term. Whenever the analysis is weaker, the possibility of closing the organization as a result of changing trends or threats will be increased. Also negligence in applying this analysis will lead to the organization's failure because it works on improving the weak areas [10].

4.1 *SWOT (Strengths, Weaknesses, Opportunities and Threats)*

The main step to figure out the role of free access in supporting the digital repositories events is through study and analysis of the vital factors of digital

repositories to identify the strengths, weaknesses, opportunities and threats elements [10]. Here we used SWOT analysis for many different digital international and Arabic repositories to learn the methods and strategies of free access and its mechanisms. The study consists of 8 repositories from international digital repositories and 2 research repositories, two digital repositories from Arabic countries, usually referred to digital repositories as digital archives or electronic editions, as they include both institutional and central repositories. These repositories include electronic copies of scientific journals articles, whether before or after the scientific arbitration, or both together. Some repositories include other types of publications, such as university theses, technical reports, electronic books, audiovisual materials ... etc. The following are models of digital repositories in the field of libraries and information science, or in other specialized fields covering this field relatively.

4.1.1 *E-LIS repository [11]*

"This repository is considered as the most significant central or objective repositories in the field of libraries and information science, related fields. It includes 6200 document in this field (until July 2007). The strength factors of this digital repository include: it supports any language (and currently owns documents published in 22 languages), its contents are considered as high-level scientific documents. More than half of the documents are accurate, many of them is systematic by nature such as university theses and conference researches. The repository adopts the classification scheme (GETA) which are developed by the work team in this digital repository. Despite the large numbers of owned documents in this repository, we cannot find presence to the Arabic countries except Lebanon (with ten documents) and Kuwait (with three documents), the repository allows free access to all of its acquisitions.

4.1.2 *Information science (DOIS) repository [12]*

This repository aims to reach the documents specialized in the field of libraries and science information, the possibility of downloading them without restrictions. And especially interested in conference researches published electronically, the repository includes 4500 conference research at the beginning of September 2007, as well as more than 15000 articles. The repository's website contains a complete list of owned periodicals titles, ordered according to the English alphabetic.

4.1.3 *Digital Library of Information Science and Technology (DLIST) repository [13]*

The digital library in information technology and its technologies is a digital repository aims to free access to specialized documents in information

sciences including fields of archive and documents management, libraries, information science, information systems, and other related fields. This repository has been developed, which was established in 2002, by both the school of information sources and libraries science, and the center of educational technology, university of Arizona. The possibility of browsing is available in the repository, also simple research and advanced research. Also can find the newly archived material by RRS technology.

4.1.4 OCLC research publications repository [14]

This repository is from the types of institutional repositories, which keen on getting information sources that are prepared or sponsored or provided by owners of OCLC center, research teams working in it, made them fully available to an audience of beneficiaries, the owned documents focused mainly on the field of libraries and information technology.

4.1.5 Librarians' Digital Library (LDL) repository [15]

Digital Library of librarians supervised by the Indian Statistical Institute, which is an objective repository, provides free access to specialized publications in the field of library and information science. It includes many types of documents, including journal articles, theses, and presentations, and images of the activities in the field of library and information science, and pictures of Indian scientist Ranganathan with possibility to browse the documents according to their themes, authors, titles or the date of publication.

4.1.6 PubMed central (BMC) repository [16]

This project was prepared and developed by the National Center for Biotechnology Information which is affiliated to the National Library of Medicine in the United States. This digital platform provides free and unrestricted benefit, and on-line, for the intellectual production of the in the medical fields, and in particular for the intellectual production published in more than 227 journal. Researchers and information specialists can reach get this Depository the intellectual production related to libraries, medical and health information.

4.1.7 Los alamos preprint depository (ARXIV) repository [17]

The server of preliminary publications in physics, mathematics, computer science, Quantum Biology and statistics. It includes more than 430,000 electronic editions of the studies in those fields and other relevant fields which are available without restrictions. Information specialists and researchers are interested in this Depository concerning the field of information retrieval, computational linguistics and information technology in general.

4.1.8 MémSIC repository [18]

It is a subjective Depository that cares for the acquisition of specialized documents in information and communication and information sciences. The website interface is in both English and French, but the main focus of the Repository on the material published is in the latter language. There is a possibility to follow the additives to the Depository right away go through RSS technology.

4.2 Researches repositories

A digital repository is a mechanism for managing and storing digital content. Repositories can be subject or institutional in their focus. Putting content into an institutional repository enables staff and institutions to manage and preserve it, and therefore derive maximum value from it. A repository can support research, learning, and administrative processes.

Repositories use open standards to ensure that the content they contain is accessible in that it can be searched and retrieved for later use. The use of these agreed international standards allows mechanisms to be set up which import, export, identify, store and retrieve the digital content within the repository.

4.2.1 Southampton university researches repository, [19]

This repository contains electronic copies of the researches, whether in the form of journal articles, book chapters, conference papers, or universities thesis and other types of research publications, including multimedia and it may also include unpublished documents and manuscripts. It also provides the full text of many of these materials free of charge.

4.2.2 IDEALS Repository [20]

The Repository of the researches of the faculty members and students of the University of Illinois at Urbana-Champaign. It provides the full text of many of these materials free of charge.

4.3 Examples of the Arabic repositories

All research resources need care and attention to survive, but digital research resources need more attention, often much sooner than resources on paper. The inherent fragility of digital materials leaves only a small window of opportunity to address this problem before we start to lose resources on an ever larger scale. Below we'll survey some of the important Arabic repositories.

4.3.1 Qatar University repository, [21]

Qatar University Institutional Repository is named Q space. It is a digital archive consists of the intel-

lectual product of the university. It manages preserves and makes the academic works available to faculty members, postgraduate students and research centers from inside the university only and in a very limited method.

4.3.2 *University of King Fahd for petrol and minerals repository, [22]*

The repository covers master's and doctoral theses licensed and registered under study, as well as the researches of university faculty members and scientific journals published by the University in addition to the electronic lectures which are available in video and electronic holdings owned by the university.

In addition to the personal data of the author and the supervisor or supervisors of the thesis, the thesis and various other materials are searched for according to substantive disciplines. It is noticeable that the Repository does not allow access to all the university theses or the different works, except for 5% only of the stock. Faculty members can view only 14 pages of the theses and the works, often the number of the pages is the title page and table of contents at the most. In addition, other visitors can only view 3 pages only of the theses or in most of the time view the abstract only. This means that access to any text almost entirely cannot be done.

The digital depository of King Fahd University of Petroleum and Minerals scored the rank 27th globally among 300 depositories. Concerning the globally institutional depository, the institutional depository of King Fahd University of Petroleum and Minerals has achieved an advanced level, in the Matrix website; it achieved the rank 23 at the global level. These successive accomplishments achieved by King Fahd University of Petroleum and Minerals University put the university on a par with the leading universities that seek to apply the optimal movement for free access. However, the Repository still imposes many restrictions on its contents and does not provide it in full form.

Hence it can be said that King Fahd University, with this effort and with its continued seeking to achieve free and full access to the university global production or to other universities, will achieve advanced ranks in this classification, this due to its follow of the best practices in the field of information availability, so that it can later achieve higher ranks and to compete on the forefront those that have a long experience in this field.

According to the above information mentioned in 4.3. It can be said that the Arab depositories in general do not support free and full access to all holdings compared to global depositories which support free and full access to all holdings either for faculty members, students or the beneficiaries without restriction, believing in the need to keep

up with the latest researches in various fields and from different institutions.

4.4 *Internal environment analysis*

Internal environmental analysis process means examining and analyzing the factors of the functions of the digital depositories and their roles in order to determine the elements of power, which is represented in the efficiency of the depositories and its ability to achieve the best achievements and elements of internal weakness, which are normally represented in the weakness of digital depositories ability. Weakness of free access operations for these depositories, or the extent of the impact of the role of the free access to the scientific production of various institutions in supporting the digital depositories which are affiliated to them so that theses digital depositories can work efficiently. Internal and external environments analysis have been applied on the digital depository of the King Fahd University of Petroleum and Minerals which is summed up in the following:

Strength elements

- The open access to information from digital repositories is a force that attracts the researchers. Because the information imported from these repositories is comprehensively valuable, meaning that it covers all areas of concerns, in addition the comprehensiveness means the presentation of information in an integrated manner clarifying different views that deal with the subject, giving a great benefit to the recipients of the information as it presents to him all different views and this is what is pursued to be achieved by the depository even if it is done partially.
- The open access to the information in the digital repository for the materials for which the free access has been applied is characterized by being free and free of charge without any restrictions that impede the use of the project.
- The digital depository is characterized by the continuous updating of information which increases another dimension to take advantage of them in all fields. This ensures the continuity of the survival of the sources and scarcity of its cease.
- Breaking the monopoly of publishers regarding the distribution of scientific research, as it makes access to scientific and technical information more just and equitable.
- Allows authors to retain the preserve the rights of publishing, and growing broadcasting for their business on a large scale.
- Accelerating the pace of scientific research and that this system allows the reduction in publishing articles from months to a few weeks or even a few days.
- Promotion of scientific productivity.

- Promotion of communication between researchers from different directions.
- Reducing the time needed for the process of scientific research.
- Facilitating the exchange and transfer of information and the possibility of conversion and transfer of data.
- Direct access to the materials available in addition to the view of the same. Keeping up with the developments right away, and then exceeding the time limits.
- The possibility of direct publishing of scientific researches, and then overcoming a number of problems, such as delaying publication of researches, and the declaration of scientific results before obsolescence.
- Follow up the scientific news such as seminars, reports, scientific activities and inventions right away.
- Upgrading and developing the status of the scientific institution that establishes the digital depository and allows free access mechanisms to the its information and sources through the increase of and intensity of view times in addition to reference citation with the intellectual production for the researchers affiliated to it in the scientific communities both locally and globally.
- The open access to information in an easy method helps greatly to benefit from the information sources that exist inside the depository, making it a permanent record for the institution intellectual, scientific and cultural life.
- Supports free access to information inside the digital depositories and makes them publicity and marketing tool for the institution that can contribute to attracting new members, students and external sources of funding and foreign grants.
- The digital depository depends on a long-term conservation of the institution intellectual production in a secure method, which facilitates access.
- Open access helps in the provision of AD services through indexing reference citations for the purpose of the qualitative and quantitative analysis to measure the performance of the researcher in the field in addition to the researcher achievement and contribution.
- Helps open access to the information to identify the institution value that establishes a depository for it which is translated into tangible benefits, represented in obtaining external sources of finance.

Weakness elements

- The heavy responsibility of sending the researches to the repositories by the academics.
- Fear of infringement of the agreements and the rights of publishers, due to lack of sufficient awareness of intellectual property rights issues.

- Technological obstacles which are represented in non-familiarity of the researchers with the skills of information free access applications.
- Lack of access in all cases to the full text.
- Lasting change in many websites addresses (URL) Uniform Resource Locator then it is probably the researcher cannot go back to get the same information or follow-up its updating.
- Control and filtration as many of the institutions, bodies or governments are still in the process of filtering or purification prior to publication, meaning publishing a part not all in all circumstances.
- Linguistic limitations, since most of the materials are available in English, which hinders the achievement of benefit to large number of those who do not speak English. In contrast, there are problems in automatic translation.
- Free access restrictions to the handicaps
- Communication problems from which many countries suffer, due to poor infrastructure.

4.5 External environment analysis

External environment analysis process helps the digital depositories in the formation of an early alarm system in order to create the necessary preparations before the emergence of possible threat within an appropriate time, and thus the design of strategies is able to meet the threat and minimize the negative effects on information free access processes in order to support the activities of digital depositories⁽¹⁰⁾ which are summed up in the following:

Opportunities

The best way to increase the impact of the information free access on digital depositories is institutional archiving, due to libraries, universities and research centers availability of intellectual products and material potentialities that qualify them for that. It is recommended that this policy shall be preceded by encouragement campaign that encourage researchers to free electronic publishing of scientific articles, whether in digital depositories, open archives or open magazines. In this regard we can mention Web Review, which can be a starting point to enter into the world of free access, as well it is necessary to involve libraries in this policy and make them aware of the possibility of using digital depositories or open archives to keep up with the current developments.

Threats

- Legal impediments related to intellectual property and copyright.
- Technological obstacles related to the infrastructure of the information technology.
- Material constraints related to financial and economic matters, financing methods and business dealings.

- Technical obstacles related to the services and standards of indexing.
- Academic obstacles related to academic upgrading systems: free access journals are not recognized by universities in the field of academic promotion for faculty members.
- Moral constraints related to scientific reputation and prestige in the publishing market.
- Access barrier handicap: Not to take into account the needs of handicaps during designing the websites of the electronic journals and digital repositories, and so the difficulty of access and benefiting this category of them.
- Language barriers: English languages is the main language in which the intellectual production available on the Internet is published, in which we find either mostly available in English or in one language only, making it difficult to those who do not speak English or another language to benefit from, especially in light of weaknesses and deficiencies of automatic electronic translation.
- Filtering and censorship barriers: The governments, institutions and researches bodies shall pick and choose what can be allowed to be made available and viewed from its scientific intellectual production, i.e. provide a part of the intellectual production not the all.
- Connectivity barriers: which are caused by the technological gap that has been able to differentiate away billions of people, including millions of serious scientists who are interested in communicating with others because of the connectivity issues that it makes that are caused by poor infrastructure for many countries.

5 ANALYTICAL VIEW

To find the proportion role of open access in support of the activities of the digital repository of the King Fahd University of Petroleum and Minerals, use the equation effectiveness of the system [10].

The effectiveness of the system

$$P_E = \frac{\text{strength}}{\text{strength} + \text{weakness}} \times 100$$

Strength points (19) while weak points (9).

$$\text{The Effectiveness } P_E = \frac{19}{19+9} \times 100 = 67\%$$

That means that open access affects 67% to support the activities of the digital repository and should be carried out to measure the effectiveness of open access on a regular basis and based upon the ineffectiveness of open access to support digital repository operations [8] = 100% – 67% = 33%.

6 CONCLUSION AND RECOMMENDATIONS

the nature of open access to information was been reviewed as well as digital repositories and the importance of open access to provide majority of benefits of digital repositories and we dealt in particular the digital repository of the King Fahd University of Petroleum and Minerals, reaching the importance of open access to influence 67% of the repository. A study (Nabil Al Sayed, 2010) that digital repositories take the lion's share of all forms of sources of open access at 80% of the total in 1001 repositories in the world due to the academic and research institutions keen to publish serious research and intellectual production of these institutions proved by study (Hanan Faraj, 2012) the low share of the Arab world in the amount of deployment in the world, which indicates that there is a knowledge gap in the Arab world is to double the content in general, and weakness in quantity and poor in quality and weakness to benefit from the information, the study proved that Egypt is the biggest Arab country—owning it owns 8 digital repositories followed by Saudi Arabia with 3 repository while Qatar, Tunisia and Sudan owns every single one reservoir only as a study showed that more Arab repository acquire many sources is the King Fahd University of Petroleum and minerals repository number 110 821 items, followed in the order of King Saud University repository number of repository 8202 items and least repository Sudanese Libraries Association number 4 items. According to a Web Matrix magazine (10) The magazine reported that 50% of Arab repository allow material full text and summaries do not provide the full text by 35.7%, and proved by study (Assyad Nabil, 2010) weak presence of the policies described operations deposit to enable open access for different materials by 57% of the total Arab repository and more deposit items readily available item not fully permissible to use the materials and the optional deposit policy in the Arab and international repositories increased by 66.6% and mandatory in 33.3% of the total global repositories hence it can be concluded the following:

- The important role played by the free access to the information stored in digital repositories in the dissemination of science and evolution.
- That the free access to information process ensures easy exchange of information between researchers.
- It also ensures free access communication between peoples and cultures, the exchange of information in digital repositories.
- Lack of Arab contributions to remember to put the odds based on the standards of developed countries.

- Upgrading and advancement of scientific institution that allows open access to their information via the mechanisms of digital repository.

Therefore, we recommend the following

- Create a national body at the level of universities, colleges and institutions meant to collect issued by academic institutions and intellectual production, archive and deposited inside the repository, such as (PubMed) of the National Institute of Health Sciences.
- Increase the highlight of free access to information movement in order to scientific content digital repositories to provide a free and quick access.
- Provide open access to research published by universities especially Arabic universities in order to enrich digital repositories.
- Interoperability and the use of metadata archiving protocols that allow indexing by search engines which increases the potential for the discovery of resources or sources stored various repository.
- To assist in the processes mail for the latest research and scientific articles publishing.
- Overcome the restrictions imposed by traditional publishing, including that high, storage and distribution and limited access to the documents alimony.
- Break the monopoly of publishers and increase the awareness of researchers to increase intellectual production taking care to note the renewed intellectual production within a short time of its issuance.

REFERENCES

- [1] Bahloul, Amina, (2014). "Open institutional archive and open access to scientific and technical information" Access date: (30, Jan, 2016). Available at: <http://www.webreview.dz/IMG/pdf/bahloul.pdf>.
- [2] Assayed Nabil Ali, (2010). "Clear Digital Arabic Content: its software and its applications and assess the needs, New York, United Nations.
- [3] Ebada Ahmed, (2011). "Digital repositories for academic institutions and their role in educational and research process" Access date: (1, Feb, 2016). Available at: <http://repository.taibahu.edu.sa/handle/123456789/4844>.
- [4] Ezzedine, Nagda, Zineb, (2014). "University libraries and initiatives to achieve free access to information and exchanging them through the electronic environment between the efforts of achievement and its obstacles" Access date: (30, Jan, 2016). Available at: <http://icoa2014.sciencesconf.org/36301>.
- [5] Ezzat, Ahmed, (2013) "Freedom of exchanging information (legal study). Access date: (30, Jan, 2016). Available at: <http://afteegypt.org/wp-content/uploads/9.pdf>.
- [6] Farah, Hanan Ahmed, (2012). "Institutional repositories of digital and its role in supporting and enriching the Arabic content on the Internet." King Fahd National Library Journal. 18. Vol 2. Access date: (1, Feb, 2016). Available at: <http://www.kfnl.org.sa/Ar/mediacenter/EMagazine/DocLib/%D8%A7%D9%84%D8%AB%D8%A7%D9%85%D9%86%20%D8%B9%D8%B4%D8%B1/93-132.pdf>.
- [7] Mahmood El Noor, Fatima, (2012). "Open access to information in university libraries (case study of the digital repository of the faculty of science in the University of Khartoum". Access date: (1, Feb, 2016). Available at: <http://khartoumspace.uofk.edu/handle/123456789/974>.
- [8] Chimezie Patrick Uzuegbu & Faustinus U. McAlber. (2012). "Digital Librarians and the Challenges of Open Access to Knowledge: The Michael Okpara University of Agriculture (MOUAU) Library Experience". Access date: (2, Feb, 2016). Available at: <http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1814&context=libphilprac>.
- [9] Kylie Pappalardo & Dr. Anne Fitzgerald. (2007). "A Guide to Developing Open Access Through Your Digital Repository". Access date: (2, Feb, 2016). Available at: <http://eprints.qut.edu.au/9671/1/9671.pdf>.
- [10] Team FME. (2013). "Swot Analysis Strategy skills". Access date: (6, Feb, 2016). Available at: <http://www.free-management-ebooks.com/dldebk-pdf/fme-swot-analysis.pdf>.
- [11] E-LIS Repository. Access date: (2, Feb, 2016). Available at: <http://eprints.rclis.org/>.
- [12] Information Science (DOIS) Repository. Access date: (6, Feb, 2016). Available at: <http://wotan.liu.edu/does/>.
- [13] Digital Library of Information Science and Technology (DLIST) Repository. Access date: (4, Feb, 2016). Available at: <http://dlist.sir.arizona.edu/>.
- [14] OCLC Research Publications Repository. Access date: (6, Feb, 2016). Available at: <http://www.oclc.org/research/publications/search.htm>
- [15] Librarians' Digital Library (LDL) Repository. Access date: (6, Feb, 2016). Available at: <https://drct.isibang.ac.in/>.
- [16] PubMed Central (BMC) Repository. Access date: (6, Feb, 2016). Available at: www.pubmedcentral.nih.gov.
- [17] Los Alamos Preprint Depository (ARXIV) Repository. Access date: (2, Feb, 2016). Available at: <http://arxiv.org>.
- [18] MémSIC Repository. Access date: (2, Feb, 2016). Available at: <http://memsic.ccsd.cnrs.fr>.
- [19] Southampton University Researches Repository. Access date: (2, Feb, 2016). Available at: <http://eprints.soton.ac.uk>.
- [20] IDEALS Repository. Access date: (4, Feb, 2016). Available at: <http://www.ideals.illinois.edu>.
- [21] Qatar University Repository. Access date: (1, Feb, 2016). Available at: <http://qspace.qu.edu.qa>.
- [22] University of King Fahd for Petrol and Minerals Repository. Access date: (6, Feb, 2016). Available at: <http://eprints.kfupm.edu.sa>.

The digital repository Arxiv: A comparative study with similar repositories

Maysa Ibrahim Yousef

Department of Information Science, King Abdul-Aziz University, Jeddah, Saudi Arabia

ABSTRACT: The domain of the free access to scientific and technical information has come into view within the past few years as an essential Development of Scientific and Technical communication, which serves the society as a whole to facilitate the Flow of information for free and without any legal or technical restrictions. It also enables the researcher to have his own Repossession without the mediation of commercial publishers and free of charge. Free access aims to achieve good communication between researchers and Institutions, meeting of the minds and scientific research and also the creation of the suitable conditions that will contribute to Advance the mechanism of scientific progress. Digital repositories are considered one of the most important mechanism of the free publishing which is used by the advocates as a Promotional tool to increase the number of digital publishing.

This sheet aims to explain the power of effective action of free access by the Digital repository Arxiv which specialized in Physics and Mathematics, declaring its role in supporting and facilitating the access of researchers to produce the Intellectual output without any legal, financial and technical restrictions. Therefore, this study adopted a comparative approach between the repository Arxiv and another four international repositories that specialized in physics and mathematics, also providing an analysis of strength to support it and weaknesses to avoid it. The study has revealed that the concerned repositories keep to create several methods for its users to retrieve various Sources of information, as well as providing free access to the majority of its material. The findings also refers that the centralized economy repository and the Physical Institute of the Russian Academy of Sciences Mathematics prove superior to the other repositories to provide free access for all users to browse without login method, it they enable reviewing the metadata of different materials with a clarity of their deposit policies. The results also show that Arxiv repository need to lay down clear deposit policies for deposit and domestic Publishing and service strategy in due form that ensures the rights of Authors, to help the beneficiaries and not to waste time waiting for communication with authors.

Keywords: Free access to information, digital repositories, Deposit policies, domestic Publishing, access policy, Arxiv repository, one-time deposit

1 INTRODUCTION

The scientific and technical information have a great importance in the information society, as it is concerned as primary and initial resource to achieve the required progress in various aspects of life because of its great benefits that help in making the right decision at the right time. It is also witnessed a great development that was widespread on the Internet, hence the digital repositories gained a great importance, especially those which are used for universities and research centers, and what it provides to save this technical and scientific information, in addition to keep the digital content of the sub organizations and their administration, and also allow meeting of the minds and expertise on a local, regional and global level. Many studies have

confirmed the importance of digital repositories, where they will form an essential part of regional blocs between libraries, universities and research institutions as one of the most successful means of Knowledge Exchange and discuss it as it works fast and it has the ability to save the information easily to protect the intellectual property of various institutions. The repositories have become so natural extension of the academic institution as a basic resource of scientific research, and also one of the most important domains in our community.

2 PREVIOUS STUDIES

There are many research and Field studies that discussed the concept of the free access, its tools and

Initiatives, although the intellectual output refers to the lack of this study at the international level, in contrast there are a large amount of theoretical studies concerning this subject which dealt with the evaluation of several repositories to identify the types of information that are available within the international and Arabian repositories, its availability and access policies. One of these pilot study is the study no. [2] Which displayed models of digital repositories in the field of international and Arab theses aiming to create Arabian digital repository for Arab theses, the study showed, according to data contained on the index of the digital repository, that the messages of the university repositories comes on the first class among the other global repositories in terms of use and allowing free access for different materials. The digital repositories for university Messages represent 51% overall coverage of the repositories that covers various sources of information, according to the statistics on the index in the year of 2010.

Study [1] has analyzed the digital repositories in the field of Librarianship and information which reached 52 repositories by using language theory parsing, using also geographically and objectivistic manner to determine the types of the sources. The study found that 71.2% of the repositories that is used in the field of Librarianship and information are from institutional repositories, and the articles were the most popular types of information sources commonly used in repositories at thereof 73.1%, followed by conferences at the rate of 57.7%, and finally unpublished materials at the rate of 53.8% of overall coverage of the repositories.

Study [4] aimed to evaluate nine repositories of the free access in the field of computer science and information technology, repositories were chosen according to the index of the free access repositories which included eight of specialized repositories in computer science in 2008 and the researcher has excluded the repositories which were designed in other languages and those which reached more than 100 document. The researcher also adopted a Questionnaire that was sent to those who are responsible of the repositories including seven elements (general information, Sources of information, content management policies, conservation policies, management audit and Feedback system. The study showed that most of the repositories were designed by one or two members of Teaching staff that determine how to retain and backup copies and how to choose all documents, the study recommended to develop the policies of conservation and content management policies and to make the full text of document obtainable.

Study [6] aims to identify digital repositories in India and evaluating it till the year of 2007, the study pointed out the current state of the Indian

digital repositories and provide a list of the best Indian digital repositories. They found that the majority of digital repositories of India allows free and full access of contents that amounting to several thousand.

Study [5] aims to release Free copy and free access to sources of education in digital repositories of India through the limitation of the institutional repositories and Periodicals of the educational free access and access records in India. it was analyzed in terms of the number of documents, used software, the growth of groups, provide the full text of the sources and types of documents in repositories and archives (Articles, university Thesis and Periodicals) the study has concluded that there are twenty-one Indian institutional repositories, and only seventeen of them made the full text of all resources obtainable while the remaining four repositories applied many restrictions on the sources.

3 RESEARCH PROBLEM

Despite the importance of research repository for open access of information and its commitment to provide full text of all resources, except that some of the free access digital repositories are still blurred and imposing many restrictions on materials for various causes.

The problem of research lies in the study of availability, free access policies, the Intellectual Property Rights of the digital repository Arxiv and what are the restrictions which are imposed on its various materials.

This study has focused on highlighting the Arxiv digital repository as one of the first digital worldwide repositories in terms of the volume of materials, unpublished theses and its various articles. The study paid a great attention to identify the repository in regard to construction, content, Techniques, access and deposit policies and the ability to make the full text of document obtainable easily and the proper management of intellectual property rights.

4 RESEARCH METHOD AND TOOLS OF THE STUDY

The comparative research was adopted by this study, which is based on a comparison of the similarities and differences between various repositories to discover the factors or circumstances that accompany the probability or practicing of certain phenomenon by comparing a single phenomenon. In this research, the free access policies of various materials among different repositories were the central Consideration of the study [3].

To achieve the main goals of this study we relied on previous studies that emphasized the importance of digital repositories, taking into consideration the similarities between the selected repositories that cover certain specialized topics in a particular area in order to figure out how to free access to materials of each repository and the methods of backup, deposit policies and the protection of intellectual property rights.

5 DIGITAL REPOSITORIES

The researchers, in the early nineties, have created the server articles Eprint server and then many researchers adopted it, they responded to the publication in the fields of physics, computer and cognitive science. Digital repositories and free periodicals have become two mechanisms of Open Access movement which has been active in the beginning at the result of individuals account who become aware of the risks and challenges facing the scientific research and scientific communication marked by the steady increase of the price of scientific journals in all areas and Budget deficit of various research institutions which led to scientific research Retirement, especially in the fields of science, technology and medicine (1). The scientific practices began to make the intellectual output available without financial constraints in the digital repositories for several years by researchers, before the movement of institutions and organizations that are involved in scientific research and the Declaration of initiatives and policies that codify the free access to information in 2002, for free, represented in hundreds of free scientific journals. Arxiv repository which.

Specialized in the field of physics, mathematics and statistics, is considered the first objective repository in the world, which was created by physicist Bo Jeans Borg as a website used for the exchange of opinion on the drafts of articles and followed by Cog. Print repository for cognitive science, languages and philosophy (7).

5.1 *Digital repository arxiv*

Digital repository was established in 1990 where Bo Jeans Borg began to send the drafts to his Office-mates via mailboxes, at this time a lot of researches were lost so he began to realize the necessary need to establish a centralized storage to keep the copies of this research. By the year of 1991a central storage was established associated with the repository of Los Alamos National Lab, which can be accessed from any computer by adding FTP.

In 1993, the repository began to allow Electronic via the Web and soon it was expanded to

include astronomy, mathematics, and computer science and quantum biology, and recently it also includes statistics by the management of both Cornell University and Simmons group and other members, and under the supervision of many universities such as the University of California, and its two Affiliates, Durham University in England and the University of Michigan and several German universities, also the Institute for Theoretical Physics in Zurich, Switzerland. These institutions enable the publishers of restoring their scientific papers and their books from the repository which aims to provide new articles and materials using Really Simple Syndication service and by subscribing to automatic notification via E-mail.

The main language of the repository is English and contains about 1124404 of publications, electronic materials and drafts also a group of unpublished theses, but it does not allow access to its contents completely and displays only the abstracts, so if you need more information of the text you must contact with the authors via e-mail provided in Abstract. It was found that the majority of beneficiaries never receive any replies or approvals to take advantage of materials, research and articles from authors leading them to search for another repository, which allows free access to various materials (7).

The objectives of the repository are as follows:

- Maintain a good number of university theses for researchers and faculty members.
- The management of these numbers of articles and theses which are not published in a proper way and provide an efficient free access to it.
- The development of several means of communication between researchers and scientists.

The website of the repository indicates that the number of articles which was added since January 2016 amounted to 1501 essays and majoring in mathematics and statistics, there are 1601 essays since the beginning of the year 2016, while the number of articles and theses on physics and that has been added from the beginning of the year 2016 amounted to 1 572 essay and thesis are available in full text, but cannot be accessed until communicating directly with authors. The number of search for the full text the same period reached 1125 search process. The website also indicates that the number of theses was about 230 thousand theses, some of them have abstracts while the other 200 thousand theses have full text but cannot be accessed.

5.1.1 *The Administrative bodies of the repository*

It is represented in the Managing Directors of various universities, members and some members of the Simmons Foundation, with some members of Supervisory universities from all over the

world who are specialized in the supervision of deposit process and supervision in general. The rest of employees who are responsible of posting the various articles, thesis and various forms of Multimedia Messages, in addition to the financial affairs that deal with all the financial affairs of the repository.

5.1.2 *The Metadata of the repository*

- Meta data of various materials

It is clear that there are special metadata of messages appears from the address and the Academic Degree as well as a summary of the study and other Attachments that are filed such as plan of the study and the personal data of the Account holder and supervisors and how to contact them by-mail.

- Searching and browsing the repository

The process of browsing E-Mails are in accordance with the objective Specializations and disciplines of the author or the year of publication, in fact this process is very easy and never waste a lot of time, also the abstracts would be obtained in different formats such as Pdf and also offers many bibliographic menus and E-mails extracts on its bibliographic form.

- Access and deposit policies of the repository

Despite the importance of free and full access to all of the different materials in a way that serves researchers and beneficiaries in various research institutions, but within the repository all members and staff are entitled to review all materials without any restrictions or conditions, while only 20% of the materials are available for the beneficiaries and are often allow a certain number of research and newly published theses containing the introduction and titles, which reduces the desired benefit of the repository hence, it would be difficult to access the full text of the messages or articles as all the results of research never allow the full texts only after communicating with the Equity owners of these materials. In addition to the Blurring of deposit policies and intellectual property rights of authors of all member institutions.

5.2 *D Space at the University of Washington*

The repository is subjected to the University of Washington, USA, it allows free access to all of its contents and the users who subscribed can participants in the repository to obtain the latest materials through using Really Simple Syndication Techniques or the notifications via e-mail. It includes about 17965 of different materials in the most important domains such as physics, mathematics, statistics and computer science, history, archeology, social sciences and other articles and

unpublished theses as well as it includes a number of databases of multimedia, it only uses the English language, noting that deposit policies are obvious at the repository (8).

5.3 *Deposit once universitäts bibliothek repository*

The repository is subjected to the Technical University in Berlin and has a lot of material which nearly about 4980 in the disciplines of earth and planetary, physics, mathematics, statistics and general technology in the form of articles, unpublished theses and drafts of works of employees of the university. The repository is enabled in German and English. The deposit policies and self-archiving introduced in a very easy and clear manner for the participants from the members of the university or outside and also allows free access to some of its materials, as it offers more than 25 pages of each material including the introduction page and it is very easy to get the summaries at the same time of the request. The repository allows depository service one time where authors can deposit their works one-time only with the ability to retrieve it later as it enables beneficiaries to search for the material according to the deposit policy that demanding obtaining permission from the Top IT Center which promote the benefits obtained by the authors to join the Deposit center and browsing its various materials, in contrast, in other deposits processes authors are not allowed to recover their books for a period of 10 years and never reuse of their Writings before the completion of the contract between the author and The repository, then they are entitled to remove their books from the depository (9).

5.4 *The centralized economy repository and the physical institute of the russian academy of sciences mathematics*

The repository is subjected to the Russian Academy of Sciences under the supervision of Russian Federation and most of the materials that are used in depository are subjected to the electoral commission of the institute. The repository allows access to the materials of metadata easily also provides full texts of research and various articles it also gives summaries of unpublished research references, conferences and many different books in addition to tables of contents of some scientific journals and owns approximately 5672 different materials in the disciplines of mathematics, statistics and physics, which are available in Russian and English.

The beneficiaries can browse the repository and access to different materials without disclosing their personal data or log in. in the case of his unwillingness to share his information either

when he logs, then he can take advantage of new materials by Really Simple Syndication techniques or notifications via e-mail or by subscribing to various forms of mailing lists. The beneficiaries can sign up for one of the activities offered by the repository or subscribe for the academic conference and through subscription they can get a list of Conference's works (10).

5.5 HAL-Rennes depository

The repository is subjected to the University of Rennes in France. It is one of the most important institutional repositories in France, adopting English and French languages. It includes approximately 13,596 different materials in general science, chemistry, physics, mathematics and statistics, astronomy, computer science, information technology, in the form of articles and business conferences in addition to published and unpublished theses and various books and patents. The number of full text theses have reached 1400 dissertation, while the rest of the materials are subjected to various policies of the process of free access to its content.

6 FREE ACCESS TO SOURCES OF INFORMATION WITHIN THE REPOSITORY AND METHODS OF SEARCH AND RECOVERY

Both of The centralized economy repository and the Physical Institute of the Russian Academy of Sciences Mathematics and Deposit Once repositories paid a great attention to the availability of the full text of sources in PDF form, while the rest of the repositories applied many restrictions on the possibility of downloading the full texts for specific sources.

The repositories made several methods of study obtainable which enables its users to retrieve various information sources including retrieval by browsing or using internal search engine and browsing according to subject or date of publication or depending to the author. This method of browsing shows all existing sources arranged according to a certain standard.

The repositories also tried to make its contents available for all to review the contents according to their subject matter and arrange information sources in a hierarchy order, while the HAL-Pennes 1 repository allows browsing the contents of the repository, according to the title of the conference or by country name and e-mail of the authors. This is the lowest forms of browsing used by the repositories, and it is considered the only repository that allows this kind of browsing.

7 DATA DESCRIPTION OF INFORMATION SOURCES AND ADDITIONAL SERVICES OF REPOSITORIES

All repositories comply with displaying Citations of the material, but The centralized economy repository and the Physical Institute of the Russian Academy of Sciences Mathematics were preeminence for showing all the bibliographic data required by the beneficiary and displaying Collection of data with a brief description of each source, as well as a detailed presentation of each source in Dublin Core format, Metadata Object Description Schema and Machine Readable Cataloging and other metadata schemes and displaying data in different forms such as Endnote, Reference Manager, Text, Excel, ASCII.

All repositories allow another services in addition to free access to some or all of their materials such as RSS feeds service that enable publishers to syndicate data automatically by notifying them via e-mails, but most of repositories do not provide external links for its various materials.

All of these repositories allow browsing of statistics for any individual, whether affiliated or non-affiliated with the possibility of preparing a profile that shows their interests and objectives to be used by their positron to facilitate the search and inform beneficiaries with the latest materials related to their Specializations through RSS or e-mail services.

8 INTELLECTUAL BUSINESS LISTING POLICIES OF REPOSITORIES

Deposit Once, D space, the central economy repository and the Institute of Mathematics adopted the way of identifying listing policies that determine the quality of the listed sources and how to insert the repository and the Necessary steps to do so, and determining the forms of files that will be included where the author can list his works in any form and then it will be updated into the proper form by the employees of The repository.

All of these repositories pointed that the majority of members of the association have the right to list works in The repository by 75%, while the digital repository Deposit Once prove superior to other repositories in providing deposit service for one-time and enables authors from outside the organization to include their intellectual production within search lists, once for all.

9 FINDINGS AND RECOMMENDATIONS

The study concluded that all repositories care to provide several methods for its users to retrieve

various information sources, as well as free access to all materials. The study also found that the centralized economy repository and the Physical Institute of the Russian Academy of Sciences Mathematics prove superior to others in providing free access for all users to browse and without logging and also allows Viewing the metadata of the materials related to materials and research summaries, as well D space repository in Washington, which gives free and full access of different materials, but at the same time it takes long time than that of a centralized economy and the Institute of Mathematics of the Russian repositories to get different materials and a clear policy of deposit, which encouraged many scientists and researchers to deposit their intellectual output.

The other repositories were unequal in the availability operations, where the HAL-Rennes 1 repository allowed 10% of the materials for all researchers and users and do not have the right to provide 25% of the materials once and for all, while the rest of the materials are accessible by different policies. For example, they provide only 14 pages of content and the introduction page is no longer one of them.

The repository Deposit once allows free access, but not for all of its contents as it allows 25 pages of each article, including the introduction page and can get these summaries at the same time of request. This repository was preeminence in using the Deposit service for one—time for the authors from within and outside the organization, which was not adopted by the others, in addition to the clearness of Deposit and self-archiving policies.

While the repository Arxiv allows researchers and beneficiaries to obtain summaries only with the possibility of communicating with authors to get various materials. Besides there are a lot of materials that the authors cannot retrieve once again as the policy of re-use is still blurring.

As we have already mentioned:

The policies of the Digital repository Arxiv should be clear in terms of new legislation and policy to browse the full text of letters, articles and published theses developing appropriate legislation that guarantee the intellectual property rights of unpublished theses drafts of articles or abstracts provided that not to include introductions. In addition to the importance of clarity of decisions that bind the employees of institutions member to

register their research as the available balance does not represent the actual balance, where the electronic publishing was estimated by 11,712 research upon 29,071 which was equivalent to 40% of researches of universities and research institutions that have participated before. And also displaying Self-archiving and free access to information between the authors and Equity owners and also clarify the benefits of free access to research and researchers and their interest if they participated in the information at the local or global level.

Increasing the Awareness of the importance of digital innovation and applying correction studies of various global repositories.

Clarifying there-use policies of different materials.

REFERENCES

- [1] Omar, Eman Fawzi. (2009). "Open digital repositories in the field of library and information". The Third Arab Forum on information technology, third-generation technologies in public libraries and information. Cairo library and information network Access date: (1, March 2016). Available on the following link: www.moltaqa.libriannet.net.
- [2] Hafez, Serfnz Ahmed Mohammed (2010). "Digital repositories of university thesis: an assessment study". Twenty-one conference of the Arab Federation for Libraries and Information, (I know). RIY-ADH: King Abdul-Aziz Public Library 491–573.
- [3] Adas, Abdul Rahman. Obidat Abdul Haq, Kayed "The concept of scientific research, its tools and methods". 17. Access date: (1, March 2016). Available on <http://www.daralfiker.com/node/6497>.
- [4] Bhat. M. H. (2009). "Open Access Repositories in Computer Science and Information Technology: an evaluation". IFLA Journal, 35, 234–257.
- [5] Das, S. G. (2007). "Open access and institutional repositories: A developing country perspective: a case study of India". IFLA Journal, 33, 229–250.
- [6] Mittal, R., & Mahesh, G. (2008). "Digital libraries and repositories in India: an evaluative study. Program: Electronic Library & Information Systems", 3, pp. 286–302.
- [7] Arxiv Repository. Access date: (Mar, 1, 2016). Available at: <http://arxiv.org/>.
- [8] D Space at the University of Washington. Access date: (Mar, 1, 2016). Available at: <https://digital.lib.washington.edu/researchworks/>.
- [9] Deposit Once. Access date: (Mar, 1, 2016). Available at: <https://depositononce.tu-berlin.de/>.
- [10] Central Economics and Mathematics Institute RAS. Access date: (Mar, 1, 2016). Available at: http://cemi.socionet.ru/oai/ecoorg_org1/oai.xml.
- [11] HAL-Rennes 1. Access date: (Mar, 1, 2016). Available at: <https://hal-univ-rennes1.archives-ouvertes.fr/>.

Pros and cons of social networks and their impact on human behaviour

A.A. Alyoubi & I.M. Alharbi

Department of Management Information Systems, College of Business, University of Jeddah, Jeddah, Saudi Arabia

ABSTRACT: Currently, no-one denies the importance of social networks, which have brought the distal and proximal closer and allow us to see people far away from us, and to share sounds and images of our countries. They also allow us to keep track of the news via our mobile phones, which is more rapid than reading newspapers or watching television. Hence, this paper aims to shed light on, and discuss in depth, the impact of social media on human psychological stress, in terms of both positive and negative effects. We conclude that for many people suffering from psychological problems it is not necessarily satisfactory to hinder them from expressing their views on social networking sites; these people find such sites to be an ideal way to share their views freely, even if they are not necessarily logical or socially acceptable. These sites have provided the possibility of catharsis of repressed feelings, regardless of their nature; but on the other hand, they allow people to isolate themselves and stop participating in social events, sporting activities and interactions with family members, particularly given the availability of social media sites on mobile phones, which helps to promote addiction. Moreover, the negative effects associated with the resulting lack of physical contact include the loss of ability to read facial expressions and body language.

1 INTRODUCTION

Currently, thanks to the information revolution, we receive a constant stream of information from around the world that cannot be regulated by standard means of border control because it is broadcast from digital devices via satellites. This has led to a race between countries to make their media more attractive by using different ways to attract viewers, including providing scientific and useful information, as well as less reliable information that allows people to ‘kill time’ and clear their minds. Today, this information stream is no longer limited to television broadcasts, because there are other news means of communicating information, and it is now up to the recipient to decide they type of information they receive by their own research and investigation. The internet, with social media as one of its various forms, reflects the information available on a particular topic as well as the level of the culture of the individual seeking or broadcasting information on that topic. Therefore, the information session, and especially the use of information in the media, has transformed the world’s media into a small village, described by Makulov (see Boediono 2011) as a ship sailing in the vast universe with its passengers being humans and other organisms. Can one be expected to go with the facts inside that ship, or small village, with continuing technological advances?

Today, social networking sites have become one of the institutions that play an important role in educating young people and equipping them with the habits and behaviours of true mission. As a tool of social change, social and educational institutions have focused on developing programmes and activities for students with a view to occupying young people’s time with activities that benefit them, as well as the intent to share aspects of mission and development, among other things, with students. The educational process is not just about teaching students, but is useful in the process of building all aspects of their personalities, inspiring a spirit of social responsibility, building self-esteem and encouraging them to assume their responsibilities in life, as well as trying to find balance in all aspects of their personalities.

According to Hampton and colleagues (2015) it makes sense to wonder if the use of digital technology creates stress. There is more information flowing into people’s lives now than ever before, much of it distressing and challenging. There are more possibilities for interruptions and distractions. It is easier now to track what friends, ‘frenemies’ and foes are doing and to monitor raises and falls in status on a near-constant basis. There is more social pressure to disclose personal information. These technologies are said to take over people’s lives, creating time and social pressures that put people at risk for the negative physical

and psychological health effects that can result from stress. Stress might come from maintaining a large network of Facebook friends, feeling jealous of their well-documented and well-appointed lives, the demands of replying to text messages, the addictive allure of photos of fantastic crafts on Pinterest, having to keep up with status updates on Twitter, and the 'fear of missing out' on activities in the lives of friends and family (Thomee 2012).

We add to this debate with a large, representative study of American adults and explore an alternative explanation for the relationship between technology use and stress. We test the possibility that a specific activity, common to many of these technologies, might be linked to stress. It is possible that technology users, especially those who use social media, are more aware of stressful events in the lives of their friends and family. This increased awareness of stressful events in other people's lives may contribute to the stress people have in their own lives. This study explores the digital-age realities of a phenomenon that is well documented: knowledge of undesirable events in other's lives carries a cost of caring (Lyons et al. 1988). Cohen and colleagues (1983) have explored the relationship between a variety of digital technology uses and psychological stress and generated a measure of stress known as the Perceived Stress Scale (PSS). The PSS consists of ten questions and measures the degree to which individuals feel that their lives are overloaded, unpredictable and uncontrollable. We used this well-established scale in the current investigation of psychological stress and the use of social media sites.

2 RELATED WORK

The use of online social networking sites to communicate with family and friends and to meet people has had a significant effect on the ways in which people interact. For some people, who are avid users of social networking sites, friendships are sustained without any face-to-face interaction. This has occurred in the past by having pen-pals, where people developed life-long friendships with people they had never met. Online social networking has increased people's capacity for making and sustaining such friendships, as well as having more regular communication with family and friends. Research is just beginning to understand the challenges, benefits and negative consequences of this different way in which people interact with each other (Australian Psychological Society 2010).

Much media attention has focused on the dangers of online social networking, particularly for young people. This has led to fears about online social networking and calls for increased regulation

and accountability of providers of these sites. Early studies suggested that internet communication had a negative impact on the individual by reducing face-to-face interactions and increasing level of loneliness (Nie & Hillygus 2002). More recent studies have revealed a more complex set of outcomes (Beer 2008). Studies have indicated that internet communications may supplement traditional social behaviour, rather than increase or decrease it. This is consistent with social network theory, which implies that the more a person socialises in a traditional sense, the more they will socialise online. Increasingly, studies are indicating that social networking has a positive impact on social connectedness and wellbeing (Valkenburg & Peter 2009). For example, a study by the Office of Communications (2008) found that those using networking sites use them to meet new people, seek out old friends, keep in touch with current friends, seek attention and/or keep up with their peers. Further, it has been suggested that online social networking may have benefits for those who find face-to-face contact difficult, such as those who are shy or introverted (Wolfradt & Doll 2001). Nevertheless, cyberbullying and inappropriate use of personal information have been identified as problems (Australian Psychological Society 2010). However, research in this area largely focuses on children and adolescents.

3 THE SPREAD OF INNOVATIONS THEORY

Rogers' (1962) theory of the spread of innovations is one of the fundamental theories of the phenomenon of the adoption of new inventions, and defines proliferation as the process by which knowledge creation or inventions spread between members of a social group or society through several communication channels. Rogers (1962) also found that there is a relationship between the spread of innovations and numerous studies of change in a society. The degree of spread of innovations depends on the effectiveness of communication within the social patterns that spread the new idea, and is also influenced by time and the stages passed through as the decision is made regarding the adoption of new technology. The first phase of this process is knowledge; where people learn of the new idea or invention and try to identify the functions of the idea or item. The second phase relates to the individual's feelings in favour of, or against, the use of innovations, while the third phase is the stage where the decision to adopt or reject the use of the innovation is made. The final stage is that in which the individual user supports the use of the innovation and emphasises its importance. During the stages

of developing an awareness of the applications of the new innovation and attempting to assess the feasibility of adopting a new technology, individuals will experiment to identify the possible benefits to be gained, even if they have reached a degree of conviction that leads to the adoption stage (Imad & Hussein). Rogers pointed out that the novelties of the adopters can be divided into five categories as follows:

Innovators: This group represents those who are eager to experiment with new ideas, and is characterised by people with a high income and higher education, openness to different cultures, and its members are less adherent to the laws of their community. They obtain their information from scientific sources and experts.

Early adopters: This group is characterised by people with greater adherence to the institutions of their group, and some of them rank opinion leaders views with their strong integration with their group.

Early majority: These are those who think twice before adopting any new ideas, and rely on their community to provide them with information. They represent a link for spreading news of the innovation to those at the centre of their society. They are located between early adopters and the late adopters.

Late majority: Rogers described this group as sceptics, since for them, embracing a new idea depends on the group they know, or they may be amenable to pressure. Members of this group are often older and have lower incomes and lower levels of education. They depend on direct connection and face-to-face contact when seeking information, rather than the media.

Laggards: This category is associated with those that value tradition and cling to old ideas, and does not embrace innovation unless it becomes outdated.

This theory is linked to the spread of several innovations and the study of factors that help to answer important research questions; for example, do you use patterns of communication and technologies that change over time? What are the attributes and characteristics of individual users of this technology? Is there a variation in the degree of its use? Relevant factors include social status, individual characteristics and attributes and users' social and cultural background. These factors affect the adoption of the use of information technology and the social dimensions associated with such use (Khaled 2009).

Based on previous studies of the spread of innovations, we are increasingly becoming aware of the importance of networks in understanding the diffusion of innovations within social patterns. Social

networks can be used to raise awareness and to persuade individuals to adopt innovations. Rogers studied communication channels, not communication tools, although other studies have pointed out the importance of addressing the diffusion process through regulatory means, as well as the influence of the social environment on the process of adopting the use of technology. The uses of entrant technologies are conditioned by a set of assumptions regarding how individuals use their means of communication and what they are seeking to achieve beyond this use, namely:

Members of the public are actors in the process of mass communication, and use the means of communication to achieve their intended goals and meet expectations.

The means of communication is used to meet the needs of the members of the public, controlling for individual differences.

Audiences choose which messages are conveyed and their content, and pilgrims compete with other sources of gratification to meet these needs.

Members of the public have the ability to determine the motives of their exposure and pilgrims choose the appropriate means to satisfy these needs.

Norms can be inferred by the prevailing cultural context, and not by their content alone.

4 HUMAN SOCIABILITY AS AN EVOLUTIONARILY STABLE STRATEGY

A steady stream of research from the complexity sciences, evolutionary psychology, biology, and neuroscience are providing a new and detailed understanding of human nature. No longer a source of armchair speculation, today's understanding of human nature is becoming a precise experimental science, drawing upon many rigorous disciplines. Not only are these findings overturning many strongly-held myths about human rationality and motivation, but they are also helping us to understand how spontaneous forms of human organisation emerge and how large-scale, self-synchronising organisations might be more effectively controlled (Clippinger 2015).

One thing that the biological sciences are demonstrating is the extent to which human beings are genetically linked to virtually all forms of life. Not only do we share 98% of our genetic code with our closest cousins, the chimpanzees, but also 46% of our genetic code with mice (Healey 2001). What is no less extraordinary is the extent to which human social behaviours are very similar to those of other social species—even those to whom we are not genetically linked (Lyons 1998). How is it that very similar cooperative strategies and social behaviours

emerge in genetically distinct species? The answer is intriguing because it argues that under certain environmental conditions, there are Evolutionarily Stable Strategies (ESSs), or circumstances where there is no incentive for any other strategy to displace the current one, that are independently discovered by different species and embedded in their respective genomes through the trial and error of thousands of generations of evolutionary testing.

What this means is that for certain forms of cooperative behaviour, there are ESSs that naturally appear as the best solutions, and that these are present in a variety of different social species: harvester ants, ravens, wolves, elephants, whales, chimpanzees and human beings. Therefore, one can argue that there are certain underlying laws—a kind of social physics—that can be abstracted for complex forms of collective behaviour and cooperation, independent of the kind of species involved. Indeed, understanding what these laws might be has been the focus of research in evolutionary game theory, multi-agent simulations and models of artificial life. The fact that highly stable strategies of collective behaviour emerge over time indicates that highly fit organisations can benefit from such strategies. However, this is only part of the picture. Human beings are unique in evolutionary history in having discovered certain survival ESSs that no other species has obtained. Therefore, it is not only important to understand how we are behaviourally similar to many other social species with similar social survival strategies, but also how we are uniquely different.

Although the architecture and functionality of the human brain and limbic systems are similar to their reptilian, mammalian and primate ancestors, there are new additions, including the neocortex, which is unique in its size and functionality. Although the human brain is composed of a large, ancient ‘legacy’ system, which like software code is patched one layer upon another without any apparent design, it is this new layer that enables a very powerful and species-specific capability.

5 SOCIAL AND PSYCHOLOGICAL ORIGINS OF USING A MEANS OF COMMUNICATION MEANS

The discovery of the relationship between the social and psychological aspects of motivation and exposure to a means of communication was made by Matilda Riley (Dannefer et al. 2005), where these aspects are dealt with in terms of:

Social aspects of the use of a means of communication: Members of the public with the means of communication are not treated as individuals

isolated from social reality, but as members of the organisation or group, and therefore demographic and social factors, such as gender, age, educational and socio-economic level, have an impact on the use of public means of communication.

Psychological aspects of the use of the means of communication: Users have certain needs to satisfy, and thus there are many incentives for them to use the media. Different individuals choose to use different media according to different psychological criteria, including their own unique psychological problems and their exposure to the means of communication as a treatment of choice for their problems.

The needs driving the use of communication media have been classified into five major categories:

Cognitive needs: These are related to information and knowledge needs and environmental monitoring.

Emotional needs: These are emotional aspects associated with needs and feelings.

Need for psychological integration: This is related to strengthening the credibility associated with self-esteem and achieving profile stability.

Need for social integration: This is linked to communication with family, friends and the world, and is based on the desire for rapprochement to others.

Escape needs: These reflect everything that is linked to entertainment and escapism.

In general, most of communication studies divide the motives for exposure and use into two categories:

Psychological motives: act to identify and target the self and the acquisition of knowledge, information and experience and all forms of education in general; reflected by the news and educational and cultural programmes.

Ritual motives: target passing time, relaxing, friendship, familiarity with the means and the escape from everyday problems; reflected by fictional programmes, such as soap operas, movies and various entertainment programmes.

It should be understood that the means of communication does not affect all individuals in all circumstances, but operates and exerts its influence in culturally and socially specific ways. Apart from satisfying functional needs and the need for leisure, exposure to specific mass communication media outlets can also satisfy needs specific to the social context in which they are used; for example, some people love to follow celebrities for both the technical or religious value of their networks.

6 PSYCHOLOGICAL STRESS BASED ON THE USE OF SOCIAL MEDIA

There is no doubt that young people are often more eager to embrace new technologies than adults, and become highly skilled very quickly. They are also more vulnerable and less inhibited in their communications than adults, and therefore may become exposed to more risks. The use of social networking sites has been the focus of a large number of research studies. The PSS (Cohen et al. 1983) consists of ten questions and measures the degree to which individuals feel that their lives are overloaded, unpredictable and uncontrollable. Participants in our study were asked if they had:

- Been upset because of something that happened unexpectedly;
- Felt that they were unable to control the important things in their life;
- Felt nervous and stressed;
- Felt confident about their ability to handle any personal problems;
- Felt that things were going their way;
- Found that they could not cope with all the things that they had to do;
- Been able to control the irritations in their life;
- Felt that they were on top of things;
- Been angered because of things that were outside of their control;
- Felt that difficulties were piling up so high that could not overcome them.

Participants responded on a four-point scale, ranging from 'frequently' to 'never'. The ten items were combined so that a higher score indicated higher psychological stress (the scale ranges from 0–30) with zero representing no stress and 30 representing the highest level.

6.1 *Relationship between psychological stress and technology use*

The respondents were asked about their use of social networking sites: the frequency with which they use different social media platforms, such as Facebook (used by 71% of internet users in this sample), Twitter (used by 18% of internet users), Instagram (17%), Pinterest (21%) and LinkedIn (22%).

Given the popularity of Facebook, we also asked very specific questions about users' networks and what people do on that platform: the number of friends (the average was 329), the frequency of status updates (the average was eight times per month), frequency of 'liking' other people's content (the average was 34 times per month), frequency of commenting (the average was 22 times per month) and how often they send private messages (the average was 15 times per month). Participants were

also asked how many digital pictures they shared online (the average was four times per week), how many people they emailed (nine people/day) and how many emails they sent and received (an average of 25 per day).

We also asked about their use of their mobile phone; the number of messages they text (an average of 32 messages per day), picture-sharing via text (an average of two pictures per day) and the number of people that they text with (an average of four people per day). Given the important differences in stress levels based on age, education, and marital and employment status, we used regression analysis to control for these factors. By using regression analysis we were able to determine the degree to which technology use is specifically associated with stress by holding demographic characteristics constant. Since men and women tend to experience stress differently, we ran separate analyses for each sex.

6.1.1 *Those who are more educated and those who are married or living with a partner report lower levels of stress*

It was found that women, particularly those with fewer years of education, tended to report higher levels of stress, while those who were married or living with a partner reported less psychological stress. For women (but not men), those who were younger, and those who were employed in paid work outside of the home also tended to experience less stress.

6.1.2 *The frequency of internet and social media use has no direct relationship to stress in men. For women, the use of some technologies is tied to lower stress*

For men, there was no relationship between psychological stress and frequent use of social media, mobile phones, or the internet more broadly. Men who used these technologies reported similar levels of stress when compared with non-users. For women, there is evidence that technology use is tied to modestly lower levels of stress. Specifically, the more pictures women shared via their mobile phones, the more emails they sent and received and the more frequently they used Twitter, the lower their reported stress. However, with the exception of Twitter, for the average person, the relationship between stress and these technologies was relatively small. Women who were heavier participants in these activities reported less stress. Compared with a woman who does not use these technologies, a woman who used Twitter several times per day, sent or received 25 emails per day and shared two digital pictures via her mobile phone per day scored 21% lower on our stress measure than a woman who did not use these technologies at all.

From this survey we were not able to definitively determine why the frequent use of some technologies was related to lower levels of reported stress for women. Existing studies have found that social sharing of both positive and negative events may be associated with emotional wellbeing and that women tend to share their emotional experiences with a wider range of people than do men (Kessler & McLeod 1984). Sharing via email, sending text messages or pictures of events shortly after they happen and expressing oneself through the small snippets of activity allowed by Twitter, may provide women with a low-demand and easily accessible coping mechanism that is not experienced, or taken advantage of, by men. It is also possible that the use of these media replaces activities or allows women to re-organise activities that would otherwise be more stressful. Previous Pew Research reports have also documented that social media users also tend to report higher levels of perceived social support. It could be that technology use leads to higher levels of perceived social support, which in turn moderates or reduces stress and subsequently reduces people's risk for the physical illnesses and psychological problems that often accompany stress (Kessler & McLeod 1984).

6.2 *Awareness of stressful events in other people's lives*

As with our analysis of psychological stress, regression analysis was used to test if the use of different digital technologies was related to higher or lower levels of awareness of stressful events in other people's lives. This allows us to determine the role of different technologies in helping different users be aware of stressful events in others' lives, controlling for likely differences in awareness that are related to demographic factors, such as age, education, race, marital and employment status.

Knowing that the sexes tend to be very different in their awareness of stressful events in the lives of those around them, we further divided our analysis into a comparison of women and men. We also anticipated that some technologies might be more commonly used for communication with those with close social ties, and primarily provide for an awareness of major events in the lives of close friends and family, while others may be more suited for awareness of events in the lives of looser acquaintances.

6.2.1 *It was found that women are more aware than men of major events in the lives of people who are close to them*

Previous research has found that women tend to be more aware of the life events of people in their social network than are men (Turner et al. 1995).

When we compared men and women based on the average number of life events that someone in their social network had experienced in the past year, women were consistently more aware than men, although the average was only statistically significant for close relationships.

6.2.2 *More educated and younger people are more aware of events in other people's lives*

A number of demographic factors were consistently related to a higher level of awareness of major events within people's social networks. For both men and women, those who were younger and those with more years of education tended to know of more major events in the lives of people around them. In addition, we found that women who were married or living with a partner, and women employed in paid work outside the home, were more aware of events in the lives of their acquaintances (weak ties), but that this was not related to awareness of events in the lives of close friends and family.

6.2.3 *Social media users are more aware of major events in the lives of people close to them*

Social media use is clearly linked to awareness of major events in other people's lives. However, the specific technologies that are associated with awareness vary for men and women.

Among both men and women, Pinterest users had a higher level of awareness of events in the lives of close friends and family. The more frequently someone used Pinterest, the more events they were aware of:

Compared with a woman who does not use Pinterest, a woman who visits Pinterest 18 days per month (average for a female Pinterest user) is typically aware of 8% more major life events from the 12 events we studied among her closest social ties.

Compared with a man who does not use Pinterest, a man who used Pinterest at a similar rate (18 days per month) would tend to be aware of 29% more major life events among their closest ties.

Men who used LinkedIn, men who sent text messages to a larger number of people, and men who commented on other people's posts more frequently on Facebook also tended to be more aware of major events in the lives of people close to them. These same technologies had no impact on woman's awareness of events in the lives of people close to them. Compared with a man with similar demographic characteristics that does not use the following technologies:

Those who send text messages to four different people via their mobile phones on an average day

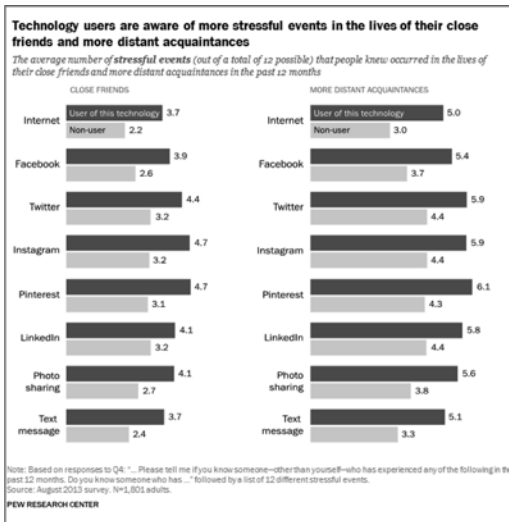


Figure 1. Various levels of stressful events based on technology.

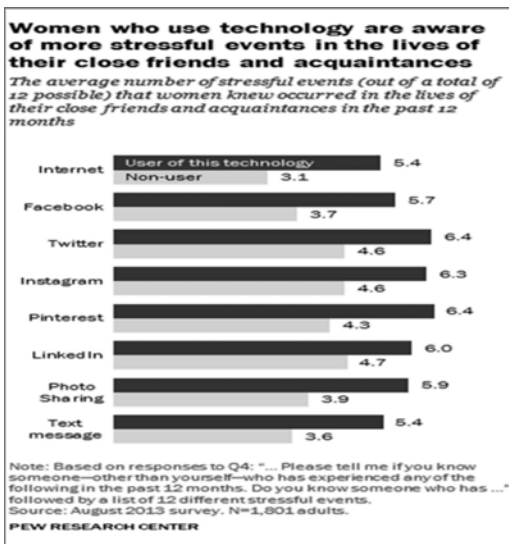


Figure 2. Various levels of stressful events for women.

(the average for a male cellphone user) tend to be aware of 16% more events among those who are close to them.

A male user of LinkedIn visits the site 15 times per month and is typically aware of 14% more events in the lives of their closest social ties.

A male Facebook user, who comments on other Facebook users' content 19 times per month, is, on average, aware of 8% more events in the lives of their closest friends and family.

For women, the more friends on their Facebook network and the more pictures they shared online per week, the more aware they were of major life events in the lives of close friends and family. Compared with demographically similar women who do not use these technologies:

A woman who shares four photos online per week tends to be aware of 7% more of the major events in the lives of those who are close to her.

A female Facebook user with 320 Facebook friends (the average for women in our sample) is, on average, aware of 13% more events in the lives of her closest social ties.

Similarly, men experienced higher levels of awareness as a result of a larger number of different technologies.

7 FUTURE TRENDS OF SOCIAL MEDIA IN OUR DAILY LIVES

An important concept related to social media and internet applications is known as the internet of things (IoT). The IoT is the people, equipment, houses, animals, means of transport and all other things linked via the internet in order to exchange information in both directions or to add value or benefits, such as monitoring or control or data collection, among other applications. In addition, it offers the possibility of making an immediate decision (real-time decisions), as shown in Figure 4. In 2014, interest in the IoT increased significantly and subsequently, much research has emerged new uses of the IoT in areas such as environment, traffic, weather and human health monitoring.

The major uses of the IoT are the following:

Smart cities: In these cities of the future everything will be linked to the positions of control systems (e.g., smart parking) across the city, and monitor the safety of buildings, bridges and infrastructure (structural health) in addition to traffic control and lighting in public roads (smart lighting) and control of solid waste, among many other applications. It is expected that by the year 2020 the number of inter-related things will reach more than 50 billion.

Smart environment: Includes forecasting and air pollution control systems.

Smart water: Includes water sources and distribution networks and leakage and consumption control.

Smart points of sale: There are self-monitoring and follow up sales operations without the presence of vendors, and provide stores with goods automatically.

Smart agriculture: These will counteract wasteful irrigation regimes, monitor soil safety quality as

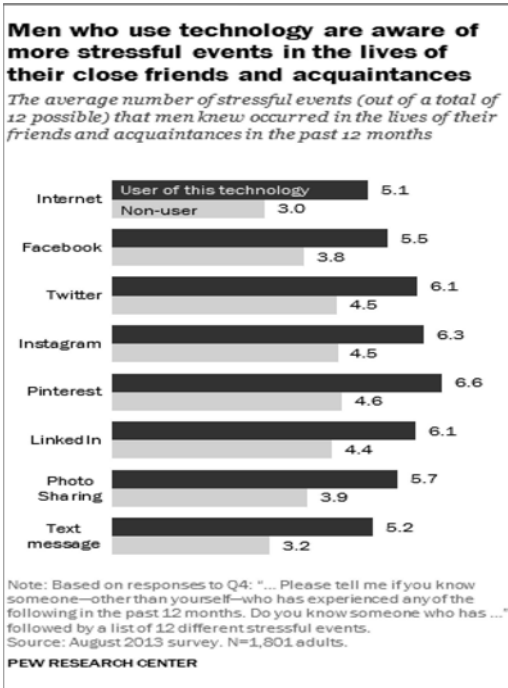


Figure 3. Various levels of stressful events for men.



Figure 4. Schematic diagram of the IoT.

well as the level of humidity and pollution control on farms.

Smart animal farms: This will be capable of observing their animals and their levels of health and hygiene, as well as trace the movements of animals.

Smart homes: These will be capable of monitoring and controlling energy and water use, as well the household appliances and the security of the property.

eHealth: These systems will monitor patients' vital signs wherever they are and on an ongoing basis, conduct laboratory tests remotely and offer potential therapeutic interventions in some cases.

Smart factories: These will automatically monitor the input and output of the manufacturing process, supply, quality control and shipping.

8 CONCLUDING COMMENTS

The biological, evolutionary, and neurological sciences are rapidly developing a rigorous scientific understanding of how people think, feel, interact, and conduct themselves as social beings. Not only will scientific knowledge replace speculation and superstition, but new forms of intervention in the genetic, cognitive, pharmaceutical, social and technological domains will greatly enhance our abilities to create more effective social organisations and institutions.

The main points related to the positive effects of online social networking as a means of reducing stress can be summarised as follows:

Optimise the benefits. If you move to a different geographic area or have less time to meet with friends in person, consider communicating with them online to maintain your friendships.

Inform yourself about security options. Talk to your friends or family about how to use these sites and read up on the options you have to keep your personal information secure.

Use your profile in a positive way. Communicate with people who have similar interests, organise social events and share information that you are comfortable with having on the internet.

Be in control of your online interactions. If you are experiencing negative interactions with someone online, stop communicating with them and consider blocking them from access to your profile.

Protect yourself. If you feel that you are being bullied by someone, think about how you can prevent that person from having access to your profile and talk to someone about what is happening.

Be respectful of others. Think carefully prior to posting personal information, including pictures or making comments about them on your site.

Do not be a bully. It is easy to make comments about others that can be hurtful or offensive. Think carefully about what you post.

Avoid going online more than you plan to. If you think you are using online networking sites too often, think about restricting yourself to a certain amount of time per day or week.

Have a process for screening people who request to be your friend. Consider the following: How well do you know them? How did they come to want to be your friend? Remember, you do not have to accept every friend request.

Be aware of the information that you post.

Never share personal information, such as your mobile phone number or address. Close friends should already have this information.

Remember that information you post online can stay there permanently, so think carefully about what you are posting.

Do not post anything online that you would normally only disclose to a close friend. When communicating online you can be drawn into providing information that you did not intend to share.

Some aspects of social networking sites are open to all members, so only post information that you are comfortable sharing with strangers in these sections.

If you really want to meet up with someone that you have only communicated with online ensure you meet in a public space, tell someone where you are going, and if possible take a friend with you.

REFERENCES

- Adler, A. 1964. *Social Interest: A Challenge to Mankind*. New York: Capricorn Books.
- Australian Psychological Society. 2010. *The social and psychological impact of online social networking*. Retrieved 13 January from <https://www.psychology.org.au/Assets/Files/Social-and-Psychological-Impact-of-Social-Networking-Sites.pdf>.
- Bediono, H. E. 2010. *Opening address to the Second International Conference on Islamic Media*. Retrieved 13 January from www.iioom.org/index.php/periodicals/downloads/?id=23.
- Beer, D. 2008. Social network(ing) sites...revisiting the story so far: A response to Danah Boyd and Nicole Ellison. *Journal of Computer-Mediated Communication* 13: 516–529.
- Clippinger, J. H. 2015. Human nature and social networks. Retrieved 13 January from https://idcubed.org/wp-content/uploads/2015/05/Human_Nature.pdf.
- Cobb, S. 1976. Social support as a mediator of life stress. *Psychosomatic Medicine* 38: 300–314.
- Cohen, S., Kamarck, T. & Mermelstein, R. 1983. A global measure of perceived stress. *Journal of Health and Social Behavior* 385–396.
- Dannefer, D. Uhlenberg, P. Foner, A. & Abeles, R. F. 2005. On the shoulders of a giant: The legacy of Matilda White Riley for gerontology. *Journal of Gerontology B* 60: S296–S304.
- Gaulin, S. J. C. & McBurney, J. H. 2000. *Social behavior. Psychology: An evolutionary approach*. Upper Saddle River, NJ: Prentice Hall.
- Hampton, K., Rainie L., Lu, W. Shin, I. & Purcell, K. 2015. *Psychological stress and social media use*. Retrieved 13 January from <http://www.pewinternet.org/2015/01/15/psychological-stress-and-social-media-use-2/>.
- Healey, J. 2001. *Genetics (Issues In Society V149)*. Spinney Press, NSW.
- Imad, H. & Hussein, L. 1998. Contact and contemporary theories, Cairo: Egyptian Lebanese House.
- Kessler, R. C. & McLeod, J. D. 1984. Sex difference in vulnerability to undesirable life events. *American Sociological Review* 49: 620–631.
- Khaled, M. bin S. 2009. *Technique of modern communication between acceptance and resistance*. Saudi Arabia: Media Department, Faculty of Arts, King Saud University.
- Lyons, R. F., Mickelson, K. D., Sullivan, M. L. & Coyne, J. 1998. Coping as a communal process. *Journal of Social and Personal Relationships* 15(5): 579–605.
- Nie, N. H. & Hillygus, D. S. 2002. The impact of internet use on sociability: Time-diary findings. *IT & Society* 1: 1–20.
- Office of Communications. 2008. *Social networking: A quantitative and qualitative research report into attitudes, behaviours and use*. Retrieved 13 January from <http://stakeholders.ofcom.org.uk/binaries/research/media-literacy/report1.pdf>
- Pennebaker, J. W., Zech, E. & Rimé, B. 2001. Disclosing and sharing emotion: Psychological, social, and health consequences. In M. S. Stroebe, R. O. Hansson, W. Stroebe & H. Schut (eds.) *Handbook of Bereavement Research: Consequences, Coping, and Care: 517–543*. Washington, DC: American Psychological Association.
- Rogers, E. M. 1962. *The diffusion of innovations*. London: Macmillan.
- Thomee, S. 2012. *ICT use and mental health in young adults*. Gothenburg, University of University of Gothenburg, Williams.
- Turner, R. J., et al. 1995. The epidemiology of social stress. *American Sociological Review* 60: 104–125.
- Valkenburg, P. M. & Peter, J. 2009. Social consequences of the internet for adolescents: A decade of research. *Current Directions in Psychological Science* 18: 1–5.
- Wolfradt, U. & Doll, J. 2001. Motives of adolescents to use the internet as a function of personality traits, personal and social factors. *Journal of Educational Computing* 24: 13–27.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Scientific and traditional communication obstacles of free access to information in institutional digital repositories

Abdulnaser S.H. Al-Msloum, Ahmed A. Al-Johani & Othman A.A. Alsulami

King Abdul-Aziz University, Jeddah, Saudi Arabia

ABSTRACT: This study aimed to identify the scientific and traditional communication Obstacles of free access to information in institutional repositories. The importance of digital repositories for universities and research centers is to raise the quality of works and scientific researches and increase the pace of research preparation in various fields and develop the community intellectually depending on beneficiary access to intellectual production to increase the scientific influence by improving mechanisms, procedures and policies that are necessary for success in services provision entrusted to the institutional digital repository in order to reach results of scientific research through publishing articles in arbitrated periodicals and Self-archiving of electronic copies. Digital repository contribute to emphasis on keeping the intellectual content and make it available for users while maintaining the rights of researchers and authors, and remove financial and legal barriers in order to access to information which lead to effective utilization of information and then provides a greater presence and impact of scientific production. Exchange ideas, make the results of scientific research available, enriching the dialogue between researchers and create proper conditions that will contribute to scientific progress. Scientific communication expresses for information exchange between specialists in direct scientific and personal contacts through scientific meetings at different levels. It could be said that the Web encouraged researchers and their institutions on acquisition the means of publication and create forms of scientific communication depends on free and complimentary access to follow completed scientific research. Obstacles to digital repository is cleared; as databases do not open for non-associated which declined scientific communication for the beneficiaries and the issue of property rights management became impediment to the beneficiaries, but it is absolutely imperative.

Keywords: digital repository, Free access, Educational establishments, Intellectual production, Information, Knowledge

1 INTRODUCTION

The world has witnessed an impressive development in terms of social and technical sciences. As the Internet emergence and communication technology development helped to transfer information to all over the world, which gave the opportunity for researchers, authors and scientists to provide scientific works and every day an essay, research, book, papers for conference or Lecture must be issued and published through specialized web sites on the internet, free access to information sites or institutional digital repositories.

Digital repositories are considered the latest digital information institutions on the Internet. These repositories have emerged in the context of free access to information initiatives; the most famous types of these repositories are “Institutional Digital Repositories” which usually follow

a university or scientific and research enterprise. These repositories make intellectual production available for those who are associated with the scientific enterprise in digital form, and then post it on the Internet to be free of charge. Institutional digital repositories play a major role in dissemination intellectual production for scientific enterprise members as they provide technical possibilities to preserve the digital content then publish and make it available for exchange information and knowledge experience at the local, regional and international level.

This paper presented the definition, importance, functions, advantages, disadvantages and characteristics of the digital institutional repository and its role in free access in order to change the concepts and theories of traditional scientific communication.

2 THE EMERGENCE OF DIGITAL REPOSITORIES

Digital repositories emerged as a result of free access movement spread through unique efforts and practices performed by researchers due to increased prices and risks facing the scientific research and communication as ARXIV repository is considered the first repository in the field of physics. Changes known to the system reflected as a new model for scientific communication since the emerge of free access movement in scientific communication in early (1999s) last century as this movement is based on the principle of re-possess the results of scientific research for researchers and research institutions through free publication online without mediation of commercial publishers⁽¹⁾. The main goal of developing digital repositories is to overcome constraints and obstacles that prevent access to information or knowledge through academic scientific publications in its traditional form all over the world, especially in developing countries⁽²⁾.

3 INSTITUTIONAL DIGITAL REPOSITORIES

(Swan) defined the institutional digital repositories as: "Digital set of outputs that have been created within a university in order to provide open access to research results in the institution⁽³⁾". While (Abad) defined the institutional digital repositories as: "Package of services provided by the institution in order to publish its digital material and intellectual production to allow access to such content online for all over the world". And also defined it as: "System for storing and archiving content and digital assets in order to be available in the process of search and retrieval later⁽⁴⁾", "Package of services provided by a university for its community members in order to manage and publish digital content created by the institution and members in the community⁽⁵⁾", "Digital collections obtained by polarization to maintain the intellectual production of one or several educational establishments combined with each other⁽⁶⁾."

From the foregoing; it is cleared that institutional digital repositories receives scientific production of society from the production, broadcast and to utilize them in all fields of human knowledge. There are significant characteristics in order to provide digital resources, hence we can define institutional digital repositories as: "Online database available on the Internet for preservation, protection and availability (in digital format) for the intellectual production of the institution and

provide appropriate physical environment for its members associated to the institution.

4 THE IMPORTANCE OF INSTITUTIONAL DIGITAL REPOSITORIES

The importance of institutional repositories increased, especially in the field of education because of the availability of the Internet and access to information resources "intellectual production of the educational institution" free of charges and legal restrictions which enables researchers to publish and share research results very easy, thus the goal of supporting free access for faculty, researchers and people interested in all various sciences would be achieved.

Scholarly Publishing and Academic Resources Coalition (SPARC) noticed that institutional digital repositories became main element of advanced infrastructure elements for scientific communications, in addition to authors and what they got of gains such as speed publishing and access to research results, as well as users who find information more easily, and the potential benefits of institutional repositories extends to the institutions which increase their research, and financiers who see expansion in the dissemination of research results. There is an international trend to publish research findings through repositories in order to maximize the impact of use for research. (Al Araby) provided a division which illustrates the importance of institutional repositories in many ways as follows⁽⁷⁾:

- a. For universities and research institutions:
Working on protecting the intellectual production of the institution, at the same time increases the institution reputation internationally through citations from the university site as a way to advertise its services. In addition; it helps in preserving the institution intellectual assets and contributing to support education, scientific research and provide the opportunity for the best use of research in the university and support educational experiences.
- b. For scientific research:
Educational institutions of universities are considered a natural source for research production and represent one of the pillars of its presence in the community as they are publishing and protect the institution intellectual production.
- c. For students and faculty:
Represents a repository where a list of researches over years is available. Since it is a resource of free access; the institutional digital repositories distribute its intellectual works widely.

- d. For the community:
Give the opportunity for free access to international researches, support knowledge exchange and provide long term preservation.

5 AIMS OF INSTITUTIONAL REPOSITORIES

Institutional digital repositories aim to adopt permanent long term preservation for the scientific institution intellectual production through publishing the associated individuals' researches whether it was private or financed by supporting scientific research centers and retrieval it more easily, in addition to saving and limiting costs incurred by the educational establishment as a result of expending huge amounts for journals in libraries and information centers and maintain scarce scientific resources.

6 ROLE OF INSTITUTIONAL DIGITAL REPOSITORIES

Institutional repositories have main functions as follows⁽⁸⁾:

- Create necessary physical environment for receiving scientific works from members associated to the institution.
- Identify procedures that must be followed in the event of handing over researches to be added in digital repository.
- Preserve scientific production for a long term, i.e. scientific production must be preserved for a long term.
- Allow free access to scientific works for beneficiaries and utilize from it.
- Maintain intellectual assets as digital repositories featured with permanent cumulative.

7 FREE ACCESS TO INTELLECTUAL PRODUCTION

Kadoura defined free access as: "Devote free access to scientific publications to overcome the ever-increasing prices of scientific journals principle, that at the economic level, while at the communication level; the principle is to share scientific information rapidly between researchers and obtain better views of scientific literature. From this perspective; the concept of free access which aims to provide information and establish a global library is permanently interchangeable⁽⁹⁾. Bjork also defined free access as: "Enabling the researcher to read a scientific research free online,

copy or publish it for non-commercial purposes free of charges and without any restrictions.

8 SCIENTIFIC AND TRADITIONAL COMMUNICATION

8.1 *Scientific communication*

Semir see that scientific communication allows scientific knowledge transfer between individuals. In case this communication was between scientists and the public it is called "scientific generalization", between specialized scientists in a particular field within closed system is called "communication between scientists⁽¹⁰⁾."

8.2 *Traditional communication*

Hurd classified traditional communication into many forms including "Informal communication": the researcher concluded arguments, meetings and visits in order to prepare report about his research, "Oral communication": the researcher identify researchers' opinions and comments regarding his research then "Direct communication" with publishers, "written communication": the researcher distribute a paper on his colleagues and other researchers around him and finally submit the research in order to be evaluated and approved before publishing by the Journal⁽¹¹⁾.

9 THE DIFFERENT BETWEEN MODERN SCIENTIFIC COMMUNICATION SYSTEM AND TRADITIONAL SYSTEM

Traditional scientific communication system contains four main groups including: researchers who produce scientific researches, publishers those in charge of publish and produce scientific journals, those in charge of collecting and posting scientific researches and users those in charge to convert researches to essential steps for discovering new knowledge. Modern scientific communication system consists of many factors; internal factors including (Researches, Publishers, Libraries specialists, Beneficiaries) and external factors including (Technology, Globalization, Economic, Governmental effects, General policies)⁽¹²⁾.

10 THE INFLUENCE OF SCIENTIFIC AND TRADITIONAL COMMUNICATION ON SCIENTIFIC PUBLICATION

10.1 *Scientific publication and traditional communication*

Scientific Publication world facing several phenomena: the crisis of providing research results costs for

scientific institutions, the concept of public property, information explosion and the web emergence with solutions related to the publication and dissemination of information. All of these factors represent incentives for the emergence of new forms of scientific communication; economically researchers complain from restrictions imposed by the publishers as the publishers are the beneficiaries of dissemination scientific research and do not make any efforts otherwise correct spelling mistakes and fix pages of printing and distribution, as well as the period of publication and broadcast research results is too long and the dissemination is not immediate which make it neglected. During the competition for publishing a large number of articles, research results are chopped which negatively affects on the results publishing.

10.2 *Scientific publication and scientific communication*

Web was established according to scientific communication model between geographically independent scientific groups, with little number but homogeneous therefore, publishing online enables scientific groups to acquire means of scientific publication, saving paper and media costs, conduct and organize contacts. Traditional scientific communication model has been replaced with new model is a modern system of scientific communication, in addition, other opportunities to broadcast and publish research works fall under a movement calling for free access to scientific research results in a form of opened digital repository rules, thus the main objective of these initiatives is to reconsider the traditional publishing systems and to encourage direct communication between researchers⁽¹⁴⁾.

11 SCIENTIFIC AND TRADITIONAL COMMUNICATION OBSTACLES IN FREE ACCESS TO INFORMATION IN INSTITUTIONAL REPOSITORIES

Bjork referred that to some obstacles which could be classified under five main axes as follows:

1. Legal obstacles related to intellectual property and authors right.
2. Technological obstacles related to information technology infrastructure.
3. Financial obstacles related to economical, financial aspects, finance methods and commercial deal.
4. Technical obstacles related to access levels.
5. Academic obstacles related to academic promotion systems in universities which did not consider free access journals in academic promotion of the faculty⁽¹⁵⁾.

12 OBSTACLES TO FREE ACCESS IN SCIENTIFIC COMMUNICATION IN DETAILS

12.1 *Legal aspects and intellectual property preservation*

Legal aspects shall be taken into account in order to preserve the rights of institution, researchers and publishers. Some legal aspects could be identified. Firstly, preserve information by copying and maintain intellectual property of the author. Secondly, not to open databases for non-associated in the institution in order to avoid the issue of intellectual property management and set out privileges for associated individuals compared to visitors knowing that, not to open databases for non-associated in the institution declined scientific communication for the beneficiaries and the issue of property rights management became impediment to the beneficiaries, but it is absolutely imperative⁽¹⁶⁾.

12.2 *Technological obstacles related to information technology infrastructure*

Some institutions place technology obstacles which declined scientific communication in digital repository including not allow to contact or transfer experience (Chat, voice call, contact via email) between user and author or publisher in the digital repository, in addition to use private server to host the repository characterized by a sense of privacy preventing personal or private advertises, non publishing communication data between the beneficiaries and associated in the digital repository, not to utilize from digital repository database⁽¹⁷⁾.

12.3 *Financial obstacles related to economical and financial aspects and finance methods*

Some institutions impose fees on individuals on joining or login digital repositories; others impose fees on requesting a digital product. These fees imposed by such institutions as the supervision body on digital repositories must ensures continuously and providing scientific services in proper way but now; such fees became an obstacle of scientific communication between beneficiaries⁽¹⁸⁾.

12.4 *Technical obstacles related to access levels*

There are several policies pursued by institutions when designing their own digital repositories, these policies limit the scientific communication activation as required, as institutions identifies and provide digital content of the institution in terms of access levels with codified terms that may

be rather complex, as well as licenses filing, ban, preservation of property rights, responsibilities and services in order to provide a high level of self-preservation of digital production offered by scientific institutions⁽¹⁹⁾.

12.5 *Academic obstacles related to academic promotion systems*

The condition of non-publishing scientific research in advance in digital repositories, as a prerequisite for scientific research provided by faculty members of scientific journals and committees for the purpose of academic promotion has declined scientific communication in many scientific share through scientific research publication in the institution's own repositories, but the associated individuals in the institution reserve full faith in the importance and necessity of research dissemination for the beneficiaries, even if the dissemination was in the account of academic promotion. It depends on the scientific institution culture in terms of cooperation, trust in each other, sense of scientific, good reputation and competitive advantage in the institution by raising its scientific prestige and the expansion in publishing works⁽²⁰⁾.

13 CONCLUSION AND RECOMMENDATIONS

In this paper, we addressed the effective role of institutional digital repositories in dissemination of explicit knowledge in various human science which adopted by scientific research centers, institutes and business organizations that prepare researches and consultation. Institutional repositories provide free access to intellectual production in order to change concepts and theories of scientific and traditional communication. There is no doubt that the communication and information technologies have affected in scientific communication pillars, therefore any community usually develop a system of values that control the behavior of associated researchers, on top of these values: scientific work became available for all society then became a part of scientific knowledge record in this area. There is no doubt that the Web encouraged researchers and their institutions on acquisition the means of publication and create models of scientific communication depends on free and complimentary access to follow completed scientific research. Scientific communication played an important role in digital repository in term of speed retrieve, free access, knowledge share between beneficiaries and individual associated to digital repository. Hence we recommend those in charge of institutional repositories to take into account activating

scientific communication in the institutional digital repositories in a proper way between beneficiaries, researchers and publishers.

REFERENCES

- [1] Al Araby, Ahmed Ebada, Digital repositories for academic institutions and their role in the educational process and research, *Journal of King Fahd Library*. Vol. 18, E. 1, Muharram—Jumada II 1433 AH, pp. 150–194.
- [2] Omar, Eman Fawzy: The emergence and development of open digital repositories—*Cybrarians Journal*—E. 27: http://journal.cybrarians.info/index.php?option=com_content&view=article&id=607:2011-12-02-01-38-43&catid=252:2011-11-28-21-19-07&Itemid=80, login date 19/02/1437.
- [3] Alma, swan, http://www.openscholarship.org/upload/docs/application/pdf/2009-01/open_access_institutional_repositories.pdf, login date 20/06/1433.
- [4] Ahmed Abad, Information and Library Network (INFLIBNET), An IUC of University Grants Commission, PB 4116, Ahmedabad-38 009, INDIA PB 4116-38 009, INDIA.
- [5] DEFINING AN INSTITUTIONAL REPOSITORY, *Library Technology Anonymous Reports* 40. 4 (Jul/Aug 2004): 6–10.
- [6] Hockx-Yu, Helen. Digital preservation in the context of institutional repositories, *Program, suppl. Institutional Repositories* 40. 3 (2006): 232–243.
- [7] Al Araby, Ahmed Ebada, Digital repositories for academic institutions and their role in the educational process and research, previous Ref.
- [8] Al Araby, Ahmed Ebada, previous Ref.
- [9] Waheed, Kadoura—Scientific communication and the free access to scientific information: Arab researchers and libraries. Tunisia: Arab League Educational, Cultural and Scientific Organization, 2006.
- [10] De Semir V. (2000). Scientific journalism: problems and perspectives *International Microbiology* 3: 125–128 http://www.upf.edu/pctstacadey/_docs/vsmicrobiology.pdf, login date 22/02/1437.
- [11] Hurd, m, julle. models of scientific communication systems. inc. 1996.
- [12] Mona Mohammed Ali El Shiekh, Digital library: Concept & Challenge. *Arab Journal of information*, Vol. 11, E 1, 2004.
- [13] Chartron, Ghislaine. Nouveaux modèles pour la communication scientifique?: “Une nouvelle donne pour les revues scientifiques?”. ENSSIB, 1997. [En ligne]. Disponible sur: <http://www.ext.upmc.fr/urfist/archives/enssib97.html> login date 05/07/1436
- [14] Salaun, Jean Michel. Publications scientifiques: Web, bibliothèques et bien public mondial. *Forum Universitaire/ communication scientifique: enjeux du partage de la connaissance*. Montréal 2000. [En ligne]. Disponible sur: http://archivesic.ccsd.cnrs.fr/action/open_file.php?url=http://archivesic.ccsd.cnrs.fr/docs/00/06/21/76/PDF/sic_00000438.pdf&docid=62176&halsid=544869383ee623ad1ff3b065c8972278, login date 06/07/14433.

- [15] Bo-Christer Bjork, A Study of Innovative Features in Scholarly Open Access Journals, 2011. <http://www.jmir.org/2011/4/e115>, login date 20/06/1436.
- [16] Legal Aspects of e-Repositories and e-Collections—JISC Legal. <http://www.jisclegal.ac.uk/Portals/12/Documents/PDFs/erepositories.pdf>, login date 23/06/1436.
- [17] Miriam A. Drake. Academic Library Challenges. (Searcher, November 2010, Vol. 18, Issue 9, pg. 17).
- [18] D-Lib Magazine. Census of Institutional Repositories in the U.S Volume 13 Number 11/12, ISSN 1082-9873, December 2007 <http://www.dlib.org/dlib/november07/rieh/11rieh.html>, login date 26/06/1436.
- [19] Peter Suber, SPARC Open Access Newsletter, issue #130, February 2, 2009. <http://www.eifl.net/faq/how-open-access-repository-policies-are-differ>, login date 01/07/1433.
- [20] 10-Ganapathi, Shinde, Deputy Librarian, Gulbarga University, Gulbarga—585 106, 2008, Karnataka, India. <http://ir.inflibnet.ac.in/dxml/bitstream/handle/1944/1146/31.pdf?sequence=1>, 05/07/1436.
- [21] Alma, swan, http://www.openscholarship.org/upload/docs/application/pdf/2009-01/open_access_institutional_repositories.pdf, login date 20/06/1436.

Assessment on course learning outcome aligned to students' achievement

Azrilah Binti AbdulAziz

Department of Information System, Faculty of Computing and Information Technology, Jeddah, Kingdom of Saudi Arabia

ABSTRACT: This study is about assessing the students' learning outcome at the end of the semester. The learning outcome are incorporated into the tests, and final examinations. The raw scores were analyzed using WinSteps Rasch analysis software which helps to treat the raw scores into probability of event. This study enables a measurement of latent traits embedded in the items to test the said ability in students. The raw scores grade, does not truly exposes the actual achievement learning outcome of the students. Therefore, a reliable instrument as the right measurement tool is crucial in providing this valuable assessment. This will ensure that the students' achievement do truly reflect the expected learning outcome in line with the industry requirement. The study reveals that the students' achieved good grades at the end of the course with different level of achievement on their expected learning outcome.

Keywords: CLO assessment; Students' achievement; Information Systems; Raw score assessment; Rasch Measurement Model

1 INTRODUCTION

One of the requirement for academic program to be accredited by the government is that it should have implemented the ABET assessment. The requirement is that students to achieve prescribed general outcomes at the end of the education program (Carnegie Mellon, n.d; Abet, 2003). Therefore, programs and courses taught within the programs should have proper planned course learning outcomes. Adding to that, there should also be some kind of mechanism to relate the achievement of the students towards those learning outcomes (Carnegie Mellon, n.d). Measurement and assessment of those learning outcome are best method of monitoring the progress of the students while providing lecturer with clearer understanding on the students' ability.

Assessment should be done with reliable measurement tool that can provide evidence and meaningful findings, so that inferences can be made. Commonly, raw scores from tests, quizzes, assignments and marks from lab activities, are used to infer on students' achievement. The higher the scores it infer the higher the students' performance. However, cumulative raw scores are only counts of discrete observations, and it is not measuring students' performance (Steven, 1946; Wright & Mok, 2013).

Therefore, valuable assessment should come from combination of measuring the learning

objectives, using reliable instrument and using the right measurement tool.

2 PROBLEM STATEMENT

Accumulated raw scores at the end of the semesters are graded and deemed as students' achievement. However, the grade may not reflect the true achievement towards the expected learning outcome prescribed for the courses. The conjunction is; a high score could probably be due to easy questions whilst a low score is contributed by very difficult questions. Therefore, some alignment of the grades and the expected learning outcomes have to be properly done, so to ensure that an overall achievement of the students ability development are met.

Until and after such prudent mechanism is resolved, students continue to fall victim to this ubiquitous unfair assessment of their true ability. The repercussion is great loss of valuable human resource to the country.

3 RESEARCH METHODOLOGY

The common process of course evaluation is that, all the portion of the marks from tests, examinations, lab activities and (or) assignments will be

summed to make up the 100% total marks at the end of the semester.

The study gathers the marks for the High Value Course Learning Outcome (HVCLO) from middle term examinations and the final exam. Even though the other course assessment included other course learning outcome, however, this study only focuses on the high value course learning outcome only.

The tests that included HVCLO are mid-term I, mid-term II and the final examination. All the marks are re-calculated to have similar base in order to avoid bias. The marks are then being analyzed using WinSteps version 3.72.1, an analysis software for Rasch Measurement model (RM).

3.1 Rasch measurement model

Rasch Measurement model (RM) is a family of Item Response Theory (IRT) that helps to explain response on observable traits to estimate unobservable traits of a construct. RM involved two important parameters in its test (1) item difficulty or the observable traits (2) respondents' ability which is the unobservable traits (Rasch, 1960/1980; Engelhard & Wind, 2013).

In Classical Test Theory (CTT) raw score or total scores are used to describe students' achievement, however RM theory uses a logistic transformation to convert raw-score observations to measures on a log-odds scale (Engelhard & Wind, 2013).

Rasch measurement model are used in this study which enable the study to probe further into the behavior pattern of the students towards the relevant items prescribed in the tests. It would help to diagnose the strength and weaknesses of both the students and those of the items.

Infit parameters in Rasch measurement model like the infit and outfit parameters are referred as "fit" statistics to indicate how accurately or predictably data fit the model. In laymen's term, it provides how close is the test to be measuring the students learning outcome. It will also help in diagnosing the ability of the students in probability of answering the tests correctly (Linacre, 2002).

3.2 The course

This course is about software quality and testing which focuses the concept of quality during the software development process. It emphasizes on the basic concepts of software quality assurance during all the stages of software development process while introduces the quality standard systems used in the field of software industry and information systems.

It is also one of the major course which the students have to take in order for them to graduate in the degree of information systems.

3.3 The sample

The students whom involved in this study are undergraduate students of a government university, comprised of 27 students from one semester. Due to some limitation, data are only collected from female students.

All the student names are hidden and only coded in increasing order with the last character representing the overall grade of achievement by the respective student. The grades will be used to compare the students' achievement with their learning outcome achievement.

3.4 Limitation of the study

The sample are taken only from one class whom enrolled in the course and was taught by one lecturer. Therefore, combination of sample between male and female may require consideration of other assessment factors.

3.5 Treatment of data

The marks from each HVCLO are only proportion to the total marks for each examination. Therefore, in order for the data not to be bias, the marks will be re-calculated so that all will have similar base marks, which is of 100.

There are five (5) HVCLO altogether and some are included in the mid-term exam and or final. The items are coded according to the availability of the HVCLO, from now onwards are referred to only CLO. If the CLO is included in mid-term 1 then it will be coded MT1CLO<number>, if the CLO is included in mid-term 2, then it will be coded as MT2CLO<number>, and if it is included in the final exam, then it will be coded FCLO<number>. The list of all the CLO are listed in Table 1 below.

Table 1. List of relevant HVCLO with their coded item names.

CLO	Coded	Remarks
CLO5	MT1CLO5, FCLO5	Able to apply statistical software quality assurance
CLO7	MT2CLO7, FCLO7	Able to define verification and validation
CLO9	MT2CLO9, FCLO9	Able to explain conventional application testing techniques
CLO10	MT2CLO10, FCLO10	Able to explain object oriented application testing techniques
CLO12	FCLO12	Able to explain web application testing techniques

The marks are entered into table in excel worksheet and then analyzed by Winsteps software to examine the psychometric properties of the items hence the test.

4 FINDINGS AND DISCUSSIONS

The summary statistics from the Winsteps analysis revealed item separation index of 0.79 with a small standard error of 0.01. The reliability index which is more than 0.7 is acceptable and considered stable when assessing different group of students it will display similar item difficulty characteristics. The maximum point of the item is at 0.06 logit and the minimum is at -0.06 logit. The location of the items are within acceptable fit parameter that is within the range of 0.5 to 1.5 for Infit MNSQ and Z-Standard (ZStd) index range is between +2.0 and -2.0 logit. Those figures are shown in Table 2 below.

The reliability index value points that the items are reasonably would be likely provide similar item difficulty if it is given to another group of students in order to test their ability in meeting the learning outcome.

However, the person reliability index for person is at 0.55 which is lower than the acceptable benchmark 0.7. This indicates that there exist inconsistency of responses among the students towards their ability in answering the items correctly. This is further revealed by the Infit MNSQ index higher than 1.5 and high index of standardized value. However, those value may be caused by one or two observations (Linacre, 2002) and their responses are less predictable (Wright & Linacre, 1994). Table 3

Table 2. Summary statistics for measured items.

Summary of 9 measured Items			
	Logit measure	INFIT MNSQ	ZStd
Maximum	0.06	1.39	1.2
Minimum	-0.06	0.46	-1.2
Item reliability			0.79
Mean item			0.00

Table 3. Summary statistics of measured person.

Summary of 27 measured person			
	Logit measure	INFIT MNSQ	ZStd
Maximum	0.29	2.35	2.0
Minimum	0.02	0.35	-1.2
Person reliability			0.55
Mean person			0.10

shows the summary statistics of measured person.

Mean person is at +0.10 logit indicating a positive students' ability on the measured learning outcome. The highest point of the student ability is located at 0.29 logit and the lowest is located at 0.02 logit. The distribution of the students' ability and the difficulty of all the items are being visualized in Figure 1 on the measurement logit ruler.

The distribution of person is observed on the left side of the vertical ruler and the items are distributed on the right side of the vertical ruler in Figure 1. The vertical ruler is the measurement ruler in measuring the persons' ability, in this study is referred to as students. The same ruler is also able to measure the difficulty of the items, in this study is referred to as CLOs.

Reading the ability of the students are like looking at an athlete on a high-jump bar; the lower the high-jump bar (the item) the easier it is for the athlete (the students) to be able to jump over it. The higher the bar (the item) the more difficulty the items are, and the more ability the athlete (the students) required.

The students have higher ability in answering the items correctly since almost all of them are located above almost all of the items (Azrilah et al., 2013). They are observed to have achieved or met almost all the expected learning outcome.

It is observed from Figure 1 that all the CLO from the final examination are found easier compared to the CLO from Mid-Term examination. All the items coded with FCLO<number> are located towards below end of the Wright map. The easiest CLO is CLO12 which is the ability to apply web application testing. The most difficult item is located at the top most location on the Wright map, MT2CLO7, which probed the students on their understanding of "verification" and "validation". It is noticed that the students find it difficult to define "verification" and "validation" earlier in the course but later find it more easier towards end of the program.

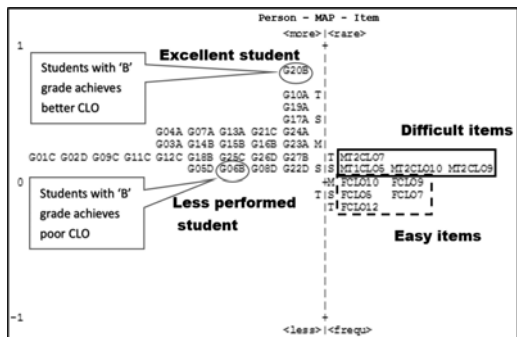


Figure 1. Distribution of person and item in Wright map.

Further observation from Figure 1 reveals that the students find the mid-term examination is little bit difficult compared to the final examination. All those items coded as MTxCLO<number> are located on top of those items coded as FCLO<number>.

On the left side of the vertical ruler, are the distribution of the person or the students. The highest person or excellent students are located on top of the Wright map, which is G20B. Further observation revealed that this student was allocated a grade B according to her summated total raw score. Commonly, it is theorized that students with high total raw score is deemed good student and vice versa. However, in this particular case, the student achieve only a “B” grade however according to their CLOs achievement, she is better than the rest of the students. Another student, G06B, whom is located towards lower end of the Wright map, is also graded as a “B” students, and however, in assessing her achievement on the learning outcome, she did not achieved the expected performance. She is among the less performing student in achieving her learning outcome.

These are the observations which can add value to the learning outcome. Assessor are required to probe further on the reason behind the findings; to answer what actually happen in that particular situation. It is crucial to know the reason why students achieve high total raw score but on other hand they do not excel in achieving the learning outcome.

5 CONCLUSION

Assessment of students are crucial in education since it will provide further knowledge towards students expected learning outcome. Commonly, assessment are done based on cumulated total raw score. However, this might not provide accurate assessment on the students’ true ability. Correct assessment tool and analysis provide better assessment and would reveal value added findings to the assessment.

This study reveals that some students with high total score, however, may not achieved or met all the learning outcome as expected. This may also happen the other way round. Therefore, it is the responsibility of the lecturer to gain further information on what are the probability of the situation, and identify corrective measure.

This study also revealed that students find it easier to achieve the expected learning outcome towards the end of the semester when all the topics

are completed. This is shown in the distribution of the CLOs on the Wright map in Figure 1, where all the items coded with FCLO<number> is located towards the bottom of the measurement ruler or the map.

Finally, it is a dangerous assumption to grade students only on their total raw score without considering their true ability. The grade that the students get at the end of the semester would determine their future. Therefore, giving correct assessment is really important so that academic institutions provide truly excellent students based on their true ability, and not by just achieving high raw marks.

6 FUTURE WORKS

Further assessment needed to be conducted to know the true issue on the different level of achievement of the expected learning outcome for the students. This would provide a framework for the next course of action should be taken by the instructor in providing quality education towards the students.

REFERENCES

- Azrilah Abdul Aziz, Mohd Saidfudin Masodi and Azami Zaharim. 2013. *Fundamental of Rasch Model: Scale Construct and Measurement Structure* (in Bahasa Malaysia). Universiti Kebangsaan Malaysia (UKM) publication.
- Carnegie Mellon n.d., *Enhancing Education*. Available from: <http://www.cmu.edu/teaching/assessment/howto/basics/grading-assessment.html>. [7 March, 2016].
- Engelhard, G. and Wind, SA. 2013. *Rating Quality Studies Using Rasch Measurement Theory*. Research Report. College Board.
- Linacre, JM. 2002. What do Infit and Outfit, Mean-square and Standardized mean?. *Rasch Measurement Transactions*, 16:2 p.878.
- Rasch, G. 1960/1980. *Probabilistic models for some intelligence and attainment test*. Copenhagen: Danish Institute for Education Research.
- Richard M. Felder and Rebecca Brent. 2003. *Journal of Engineering Education*, 92(1): 7–25.
- Stevens, S.S. (1946). *On the Theory of Scales of Measurement*. *Science*, 103, 667–680.
- Wright, BD. and Linacre, JM. 1994. Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8:3 p.370.
- Wright, BD. and Mok. 2013. *An Overview of the Family of Rasch Measurement Models*, a book chapter in *Introduction to Rasch Measurement*. Downloaded from www.jampress.org/irmch1.pdf on 18 March, 2013.

Management aspects of big data in various enterprises

Bader A. Alyoubi

Department of Management Information Systems, College of Business, University of Jeddah, Jeddah, Saudi Arabia

Ibrahiem M.M. El Emary

Department of Information Science, King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT: Currently, big data exists all around us, with its applications including a wide variety of sectors, from health and education to financial markets, among others. Big data can assist in providing a competitive advantage for all organisations if it is managed well and it returns benefits from its analysis. This is because it provides a deeper understanding of an organisation's customers and their requirements. It also helps ensure that decision-making within an organisation is more effective based on the information retrieved from customer databases; thus increasing efficiency and profit, while reducing losses. Data can be derived from various sources inside or outside the organisation to add significant value to multiple functions and purposes, provided that there is transparency and good management. Moreover, data collected for a particular purpose can be re-used for other purposes, reducing the need for data that may be expensive and time-consuming to collect. Accordingly, this paper aims to address and discuss the management aspects of big data in various organisations that hope to achieve significant benefits at all business levels. It is expected that large savings can be made in both time and money when institutions and government agencies manage their sensitive datasets effectively and analyse their big data requirements and practices.

1 INTRODUCTION

Big data is a term covering several aspects, ranging from a technology base to a set of economic models. In this paper, the following definition of big data will be applied: 'big data' is a term encompassing the use of techniques to capture, process, analyse and visualise potentially large datasets in a reasonable timeframe that is not accessible to standard Information Technologies (IT) (NESSI 2012). By extension, the platform, tools and software used for this purpose are collectively called 'big data technologies'.

Big data is not a new concept, and can be seen as a moving target linked to a technology context. The new aspect of big data lies within the economic cost of storing and processing such datasets; the unit cost of storage has decreased by many orders of magnitude, amplified by the Cloud business model, significantly lowering the upfront IT investment costs for all businesses. As a consequence, 'big data concerns' have moved from big businesses and state research centres to mainstream status (NESSI 2012).

On other hand, big data is a term referring to datasets or combinations of datasets whose size (volume), complexity (variability) and rate of

growth (velocity) make them difficult to be captured, managed, processed or analysed by conventional technologies and tools, such as relational databases and desktop statistics or visualisation packages, within the timeframe necessary to make them useful. While the size used to determine whether a particular dataset is considered big data is not firmly defined and continues to change over time, most analysts and practitioners currently refer to datasets from 30–50 terabytes (1,012 or 1,000 gigabytes per terabyte) to multiple petabytes (1,015 or 1,000 terabytes per petabyte) as big data (NAVINT Partners 2012).

On any day, more data will be produced than the amount of information contained in all printed material in the world. The Internet Data Center estimated the rate of growth of data to be of a factor of 300 between 2005 and 2020, expected to rise from 130 exabytes to 20,000 exabytes (Gantz & Reinsel 2012). The complex nature of big data is primarily driven by the unstructured nature of much of the data that is generated by modern technologies, such as that from web logs, radio frequency identification (RFID), sensors embedded in devices, machinery, vehicles, Internet searches, social networks such as Facebook, portable computers, smart phones and other cell phones, GPS devices and call centre records.

In most cases, in order to effectively utilise big data, it must be combined with structured data (typically from a relational database) from a more conventional business application, such as Enterprise Resource Planning (ERP) or Customer Relationship Management (CRM). Similar to the complexity, or variability, aspect of big data, its rate of growth, or velocity aspect, is largely due to the ubiquitous nature of modern online, real-time data capture devices, systems and networks. It is expected that the rate of growth of big data will continue to increase for the foreseeable future.

Specific new big data technologies and tools have been, and continue to be, developed. Much of the new big data technology relies heavily on massively parallel processing databases, which can concurrently distribute the processing of very large sets of data across many servers. When big data is effectively and efficiently captured, processed, managed and analysed, companies are able to gain a more complete understanding of their business, customers, products and competitors, which can lead to efficiency improvements, increased sales, lower costs, better customer service and/or improved products and services.

2 MAJOR CHALLENGES FACING BIG DATA IN DIFFERENT ORGANISATIONS

The effective use of big data has the potential to transform economies, delivering a new wave of productivity growth and consumer surplus. Using big data will become a key basis of competition for existing companies, and will create new competitors who are able to attract employees that have the critical skills for a big data world (Almeida, 2013). In the following, we will address the major challenges facing big data:

Understanding and utilising big data: It is a daunting task in most industries and companies that deal with big data to understand the data that is available to be used and to determine the best use of that data based on the company's industry, strategy and tactics. Also, these types of analyses need to be performed on an ongoing basis as the data landscape changes at an ever-increasing rate, and as executives develop more and more of an appetite for analytics based on all available information.

New, complex, and continuously emerging technologies: Since much of the technology that is required in order to utilise big data is new to most organisations, it will be necessary for them to learn about these new technologies at an ever-accelerating pace, and potentially engage with different technology providers and partners than they have used in the past. As with all technology, firms entering into the world of big data will need

to balance the business needs associated with big data with the associated costs of entering into, and remaining engaged in, big data capture, storage, processing, and analysis.

Cloud-based solutions: A new class of business software applications has emerged whereby company data is managed and stored in data centres around the globe. While these solutions range from ERP, CRM, document management, data warehouses and business intelligence to many others, the common issue remains the safekeeping and management of confidential company data. These solutions often offer companies tremendous flexibility and cost-saving opportunities compared to more traditional on-premise solutions, but they raises a new dimension related to data security and the overall management of an enterprise's big data paradigm.

Privacy, security, and regulatory considerations: Given the volume and complexity of big data, it is challenging for most firms to obtain a reliable grasp of the content of all of their data and to capture and secure it adequately, so that confidential and/or private business and customer data are not accessed by and/or disclosed to unauthorised parties. The costs of a data privacy breach can be enormous. For instance, in the healthcare field, class action lawsuits have been filed where the plaintiff has sought \$1,000 per patient record that has been inappropriately accessed or lost.

In the regulatory area, for instance, the proper storage and transmission of personally identifiable information, including that contained in unstructured data such as emails, can be problematic and necessitate new and improved security measures and technologies. For companies doing business globally there are significant differences in privacy laws between the US and other countries. Lastly, it is very important for most firms to tightly integrate their big data, data security/privacy and regulatory functions.

Archiving and disposal of big data: Since big data will lose its value for current decision-making over time, and since it is voluminous and varied in content and structure, it is necessary to utilise new tools, technologies and methods to archive and delete big data, without sacrificing the effectiveness of using big data for current business needs.

The need for IT, data analysts, and management resources: It is estimated that there is a need for approximately 140,000–190,000 more workers with deep analytical expertise and 1.5 million more data-literate managers, either re-trained or newly hired. Therefore, it is likely that any firm that undertakes a big data initiative will need to either re-train existing people, or engage new people in order for their initiative to be successful.

Leaders of organisations need to recognise the potential opportunities as well as the strategic threats that big data represents and should assess

and then close any gap between their current IT capabilities and their data strategy and what is necessary to capture big data opportunities relevant to their enterprise. In this task, they will need to be creative and proactive in determining which pools of data they can combine to create value and how to gain access to these data pools.

3 BIG DATA SOURCES

Big data analytics has started to impact all types of organisations, as it carries the potential power to extract embedded knowledge from large volumes of data and react accordingly in real time. We exemplify some of the benefits by exploring the following different scenarios. New technologies produce massive streams of data in real time and space that with time can make it possible to extract patterns of how the structure and form of a city changes and the way in which its citizens behave. In these 'smart cities', data gathered by sensors can be integrated with transport data, financial transactions, the location of users and social network interactions, and will provide an entirely new dimension to thinking about how cities function. The dangers associated with this aspect relate to privacy and will be elaborated on below. Managing data in an effective way opens a wide field of opportunities for cities contributing to the improvement of services for citizens, such as 'on-demand' and context-sensitive transportation strategies, optimised management of energy demands, more holistic and preventive healthcare approaches and the development of new services, such as e-voting, among others.

Various branches of experimental science generate vast volumes of experimental data. Petabytes of data per day is not uncommon in these fields (e.g., research in particle physics produces vast amounts of experimental data within short time frames). Fulfilling the demands of science requires a new way of handling data. Optimal solutions provided by big data technologies to analyse and properly compare disperse and huge datasets would provide huge benefits in terms of discoveries in experimental sciences.

Big data in healthcare is associated with an explosive increase in the volume of patient-specific data. A prime example is medical imaging, where even small pathological features measuring just a few millimetres can be detected by magnetic resonance imaging and on computed tomography scans. Doctors, already under significant time and cost pressure, will find it increasingly difficult to conduct their own analyses and error-free evaluations of the growing volume of data and images. They urgently need the support of automated solutions, which are increasingly based on the

application of machine learning to large sets of example images. An even more dramatic increase in data volume is expected with the introduction of modern molecular analysis to medical practice; for example, by the increasing use of next-generation DNA sequencing. Naturally, imaging data and molecular data need to be evaluated in the context of complete patient information, including all clinical data as well as family history. Data protection and information security are particularly sensitive issues when it comes to medical data. If there is even the slightest perception that health records could fall into the wrong hands, the corresponding services will not be adopted. The amount of mobile data traffic is expected to grow to 10.8 exabytes per month by 2016. This tremendous growth is driven mainly by the increased usage of smart phones and tablets.

Big data technology is needed in order to realise some advanced use cases in today's mobile networks and will be certainly required in future networks. Big data is important for example for managing and operating mobile networks and gaining insights into the network with the goal to improve network quality; which includes the isolation and correction of faults within the network, support of security-related detection and prevention mechanisms, traffic planning, prediction of hardware maintenance, or the calculation of drop call probability.

The changes brought by new social media technologies mainly refer to the appearance of new types of content providers and new types of content, often referred to as 'new media'. These new media are giving the power of speech to citizens, who can now very easily report, blog and send short text messages (e.g., tweets), rapidly creating huge amounts of new content. Traditionally, in the area of news media, conventional journalism has been the main trend, operating with standard news collection and broadcasting procedures while mediating mainstream types of content (e.g., politics, sport, economy, culture, health) from authoritative sources. However, over the last few years new Internet technologies have appeared and have disrupted this business process. Traditional news media are increasingly being overtaken by the rise of web news services. In terms of the associated business and economic activities, many software and service vendors already rely on online analytical programming systems to perform their market or sales analysis. Big data technologies do not provide a clear advantage here, and can be at best viewed as enablers to help scale legacy systems. However, in order to move beyond this state and access finer details, past simple statistics, different approaches that are best supported by big data technologies are required.

Big data is also a wrapper for different types of granular data. Below, we list five key sources of high volume data: (1) public data, (2) private data, (3) data exhaust, (4) community data and (5) self-quantification data. ‘Public data’ are data typically held by governments, government organisations and local communities that can potentially be harnessed for wide-ranging business and management applications. Examples of such data include those concerning transportation, energy use, and healthcare that are accessed under certain restrictions in order to guard individual privacy. ‘Private data’ are data held by private firms, non-profit organisations and individuals, which reflect private information that cannot readily be imputed from public sources. For example, private data include consumer transactions, RFID tags used by organisational supply chains, movement of company goods and resources, website browsing and mobile phone usage, among several others. ‘Data exhaust’ refers to ambient data that are passively collected, non-core data with limited or zero value to the original data-collection partner. These data were collected for a different purpose, but can be recombined with other data sources to create new sources of value. When individuals adopt and use new technologies (e.g., mobile phones), they generate ambient data as by-products of their everyday activities. Individuals may also be passively emitting information as they go about their daily lives (e.g., when they make purchases, even at informal markets; when they access basic healthcare or when they interact with others) (George, Haas & Pentland 2014). Another source of data exhaust is information-seeking behaviour, which can be used to infer people’s needs, desires or intentions. This includes Internet searches and calls to telephone hotlines, or other types of private call centres. ‘Community data’ is a distillation of unstructured data—especially text—into dynamic networks that capture social trends. Typical community data include consumer reviews on products, voting buttons (e.g., ‘I find this review useful’), and Twitter feeds, among many others. These community data can then be distilled for meaning to infer patterns in social structure [e.g., Kennedy (2008)]. ‘Self-quantification data’ are types of data that are revealed by the individual through quantifying personal actions and behaviours. For example, a common form of self-quantification data is that obtained through the wristbands that monitor exercise and movement, generating data that are then uploaded to a mobile phone application and then tracked and aggregated. In psychology, individuals have ‘stated preferences’ of what they would like to do, versus ‘revealed preferences’, wherein the preference for an action or behaviour is inferred. For example, an individual might buy

energy-efficient light bulbs with the goal of saving electricity, but, instead, keep the lights on longer because they are now using less energy. Such self-quantification data helps bridge the connection between psychology and behaviour. Social science scholars from diverse areas, such as psychology, marketing and public policy, could benefit from stated and implicit preference data for use in their research.

4 DATA MANAGEMENT PERFORMANCE

Performance and scalability are central technical issues necessary to deal with the huge volume of data to be stored and processed by big data systems and technologies. Two primary groups of technical issues call for significant advancements and industrially applicable research results. On the one hand, there is a need for novel effective solutions dealing with the issue of data volume per se, in order to enable the feasible, cost-effective and scalable storage and processing of enormous quantities of data (NESSI 2012). Promising areas that call for further investigation and industrially applicable results include effective non-uniform replication, selective multi-level caching, advanced techniques for distributed indexing and distributed parallel processing over data subsets with consistent merging of partial results, with no need for strict consistency at any time.

On the other hand, another relevant performance/scalability issue worth significant effort relates to the need that big data analysis be performed within time constraints, as required in several application domains. The possibility to define quality constraints on both big data storage (e.g., where, with which degree of replication, and with which latency requirements) and processing (e.g., where, with which parallelism, and with which requirements for computing resources) should be carefully taken into account (NESSI 2012). Currently, process analysis in areas such as business process management is disconnected from database analyses in areas such as data mining and business intelligence. Bridges focusing on process mining will be also essential for progression of the big data theme.

A further perspective in conceiving of a big data business ecosystem is that of big data market places. Despite the value that is hidden in big datasets, there are certain challenges associated with the cost and complexity of publishing such data, as well as the cost and complexity of consuming and utilising it. Many data providers who have stockpiled vast amounts of interesting data struggle with the problem of finding ideas for creating novel services using their data, identifying

what makes their data relevant for potential consumers, and deploying solutions for rapid integration of data for loosely defined services. The lack of a sustainable data marketplace ecosystem for big data, where producers of data can effectively disseminate their data and consumers can interact with the providers in new ways, enabling efficient delivery of new applications and services, hinders the development of novel data-driven business models in the big data domain. Current data markets face various problems, such as data discovery, curating, linking, synchronisation and distribution, business modelling, sales and marketing. This situation calls for new technologies, infrastructure, approaches and methodologies to create and sustain a big data marketplace ecosystem (NESSI 2012). This will require, among others aspects, improving current practices for publishing and consuming big data; tool-supported methodologies for efficient publication, dissemination and consumption of data in data marketplaces; scalable data dissemination and communication approaches between data providers and consumers in the data marketplaces.

Because the current technology enables us to efficiently store and query large datasets, the focus is now on techniques that make use of the complete dataset, instead of sampling. This has tremendous implications in areas like machine learning, pattern recognition and classification, to name a few. Therefore, there are a number of requirements for moving beyond standard data mining techniques:

- A solid scientific foundation to be able to select an adequate method or design;
- New algorithms (and proof of their efficiency and scalability, etc.);
- A technology platform and adequate development skills to be able to implement it;
- A genuine ability to understand not only the data structure (and the usability for a given processing method), but also the business value.

As a result, building multi-disciplinary teams of 'data scientists' is often an essential means of gaining a competitive edge. More than ever, intellectual property and patent portfolios are becoming essential assets. One of the obstacles to the widespread adoption of analytics is a lack of understanding of how to use analytics to improve a business. The objects to be modelled and simulated are complex and massive, and correspondingly, the data is vast and distributed. At the same time, the modelling and simulation software solutions are expected to be simple and general, built on solid foundations provided by a few robust computational paradigms and naturally oriented towards distributed and parallel computing. Hence, new methodologies and tools for data visualisation and simulation are required.

Due to the increasing mobility of users and devices, context awareness is increasing in importance. A suitable and efficient content—and context-aware routing of data is needed in many cases. Facing existing infrastructure and big data set-ups, many solutions focus on processing and routing all data at once. For example, in manufacturing existing data has no relation to the context of the user's history, location, tasks, habits and schedule. Concepts for taking the spatial users into account are a major challenge. The goal is to take the context into account for data that is not related to a user or context and present the right data to the right people and devices.

Data visualisation is vital if people are to consume big data effectively. The reports generated from the analytics can be thought of as documents. These documents frequently contain varying forms of media in addition to textual representation. Even if textual representation alone is used, the sheer amount in large and complex documents requires carefully designed presentation for a digital screen. When trying to represent complex information and the associated rationale, tasks, social networks and conceptual networks on screen(s), the design issues multiply rapidly. The interface for such information needs to be humane, i.e., responsive to human needs and considerate of human frailties. The frailties and needs of knowledge workers are closely linked. They need relevant information in a just-in-time manner; but too much information, which they cannot search efficiently, can hide that which is most relevant. They need to understand the relevance and relatedness of information, but frequently have other work commitments that stop them from striving to establish relevance and relatedness.

5 DATA SHARING, PRIVACY AND ETHICS

In current IT infrastructures, the provision of services, such as network connectivity, is usually associated with a Service Level Agreement (SLA) defining the nature and quality of the service to be provided. Such SLAs are important to limit liability, to enable better provisioning of the operational infrastructure for the provider and to provide a framework for differential pricing. The exponential expansion of network connectivity and web services was, in large part, due to significant technological advances in the automation of SLA enforcement, in terms of monitoring and verification of compliance with the contract. In contrast, the realm of big data sharing agreements remains informal, poorly structured, manually enforced and linked to isolated transactions (Koutroumpis & Leiponen 2013). This acts as a significant barrier

to the market in data—especially for social science and management research, which cannot access these private data for integration with other public sources. Data sharing agreements need to be linked into the mechanisms for data protection and privacy, including anonymisation for open data, access control, rights management, and data usage control. Issues such as imputed identity, where individual identity can be inferred through data triangulation from multiple sources, will need to be carefully considered and explicitly acknowledged and permitted. Management scholars will be invited to embed themselves into social issues based on defining research questions that integrate data sharing and privacy as part of their research methodology. Doing so will likely allow us to refine the model for data sharing and data rights, which could be universally beneficial and define big data collaborations in the future.

6 CONCLUDING REMARKS

Big data is of economic and scientific importance. It is a scientific belief that the bigger the data used in a research, the better the accuracy. Data are created every second in real life which means the volume of data available can never reduce. In fact, IDC's Digital Universe study predicts that between 2009 and 2020, digital data will grow 44-fold to 35 zettabytes per year. It is also important to recognise that much of this data explosion is the result of an explosion in devices located at the periphery of the network, including embedded sensors, smart phones and tablet computers. The challenges of big data lie within the organisation, as well as in the environment. Regulatory laws and privacy concerns will have an effect on how data will be used and may limit the usage of data from a business perspective to protect the private individual from a breach of privacy. When misused, big data has the potential to threaten the privacy of innumerable individuals. As society must redefine its understanding of a society with big data technology,

a regulatory framework is necessary make sure that the use of big data stays within boundaries. A new ethical code needs to be developed that defines the use of big data technology from an ethical point of view. Within an organisation, big data must justify its purpose as management tool and challenges the corporate culture to redefine management decision-making. Depending on the corporate culture, big data will find its way sooner or later into management. Smaller businesses are generally more flexible and adapt more easily to new technologies and new ways of working.

REFERENCES

- Almeida, F. 2013. The main challenges and issues of big data management. *International Journal of Research Studies in Computing* 2(1):11–20.
- Gantz, G. & Reinsel, D. 2012. *The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the Far East*. Technical Report from the Internet Data Center.
- George, G., Hass, M. R. & Pentland, A. 2014. From the editors: Big data and management. *Academy of Management Journal* 57(2):321–326.
- Kennedy, M. T. 2008. Getting counted: Markets, media, and reality. *American Sociological Review* 73:270–295.
- Koutroumpis, P. & Leiponen, A. 2013. Understanding the value of (big) data. In *Proceedings of 2013 IEEE International Conference on Big Data*. 38–42. Silicon Valley, CA, 6–9 October 2013.
- McKinsey Global Institute. 2011. *Big data: The next frontier for innovation, competition, and productivity*. Los Alamitos, CA: IEEE Computer Society Press.
- Munford, M. 2014. *Rule changes and big data revolutionize Caterham F1 chances*. The Telegraph, Technology Section, 23 February 2014.
- NAVINT Partners. 2012. *Why is BIG data important?* Retrieved 11 January from <http://www.navint.com/images/Big.Data.pdf>.
- NESSI. 2012. *Big data: A new world of opportunities*. Retrieved 11 January from http://www.nessi-europe.eu/Files/Private/NESSI_WhitePaper_BigData.pdf.
- Pentland, A. 2014. *Social physics*. New York, NY: Penguin.
- Wilson, E. O. 1998. *Consilience: The unity of knowledge*. New York, NY: Knopf.

Towards an approach based on hadoop to improve and organize online search results in big data environment

K. Aoulad Abdelouarit

*Information Technology and Modeling Systems Research Unit
Computer Science, Operational Research and Applied Statistics Laboratory
Abdelmalek Essaadi University, Tetuan, Morocco*

B. Sbihi

*Information Technology and Modeling Systems Research Unit
Computer Science, Operational Research and Applied Statistics Laboratory
I-School ESI, Rabat, Morocco*

N. Aknin

*Information Technology and Modeling Systems Research Unit
Computer Science, Operational Research and Applied Statistics Laboratory
Abdelmalek Essaadi University, Tetuan, Morocco*

ABSTRACT: In this article we study the technical specifications required for the proper conduct of online search process in Big Data environment, with the intention to evaluate the consistency of collected data and identify opportunities to improve and organize search results through a fictitious model that can make them well presentable and their information easily consumable in the future. The online data volume has increased dramatically but the quality of the information brought in these data and their form of presentation has clearly deteriorated. This is mainly due to the problem of representation the majority of the data generated in the Web in an unstructured form of information. Which prevents traditional search engines to effectively meet the information needs expressed by users or applications. It is in this context that we propose to design a technique that process massive and unstructured data to improve and organize online search results. Our solution is based on the combination of three systems: Hadoop, Lucene and Solr. As a result of this solution, massive and unstructured data can be processed from Big Data layer to structure them by Hadoop technique, in addition to index them by Lucene engine and finally organize their information to be accessible for online search through Solr framework.

Keywords: Big Data, Online Search, Unstructured Data, Hadoop

1 INTRODUCTION

With the emergence of the Web 2.0, a new vision of the Web was created by considering the user as a potential producer of information and not just a consumer (Sbihi et al. 2010). This radical change has significantly increased the amount of Internet data known as Big Data. The data of Big Data phenomenon represent the largest portion of data on the Internet. This mass of data that occupies our daily life does not cease to increase and requires advanced ways to capture, communicate, aggregate, store and analyze (Matei 2014).

Blogs, social networks, wikis, etc., are one of the reason of the large amount of data in Internet. This impacts directly the online search systems:

where everyone is launching its query to obtain a particular result, but since data comes from multiple sources, the result becomes large and rich (Gayathri et al. 2013). The Big Data phenomenon made possible the development of highly skilled online search engines. The web pages generated by search engines are based on search terms that require sophisticated algorithms and ability to handle a huge number of requests (Lakhani et al. 2015). In their Internet search, users often rely on the best elements of the results page returned by search engines, and just beyond the first few pages, although the summaries of search findings are less relevant than other results on other pages. Currently, search engines offer automated queries to assist users in their information search, but the

proposed requests are often out of context and based on popular searches, rather than on the specific information needs of the user (Leeder et al. 2016). This leads us to ask: how do we customize search results such a way that users can gain more profit while online searching for information?

On the other hand, massive data returned by online search engines are not necessarily textual and do not always manifest itself in a structured format. What makes them difficult to use by users and applications and benefit from the wealth of its information (Aoulad Abdelouarit et al. 2015). So we can make our question more specific: how do we deal with massive and unstructured data returned by the search engines to improve their presentation and consumption by users? Also, is there any concrete model to represent this kind of data that are not necessarily in text format?

This article is the following of the problem of massiveness and unstructured data generated from Big Data phenomenon that we have already exposed in our previous work (Aoulad Abdelouarit et al. 2015). As we said, this problem impacts negatively the online search process. Thus, we study the technical specifications required to the success of this process in the Big Data environment. Our solution is based on the integration of three systems:

- a system for processing massive and heterogeneous data generated from the Big Data layer;
- a framework that offers searching on massive data;
- a tool as an indexing engine for data.

The following paragraph presents the state of the online search in Big Data environment by exposing the problem of the massive and unstructured data of the Web and its impact on online search results; the paragraph 3 presents the concepts and approaches of existing solutions to process the massive and unstructured data. Then we present in paragraph 4 the used solution based on the Hadoop technique integrated with the Solr search system, which tends towards the customization of search results for a better organization and improvement of use. The last paragraph presents a general conclusion putting forward a series of perspectives.

2 THE ONLINE SEARCH IN BIG DATA ENVIRONMENT

2.1 Executing the online search process in big data environment

The scope and volume of information on the Web requires good search skills such as the ability to formulate relevant keywords to find the

information. However, most users are unable to limit the subjects of their search and are overwhelmed by the amount of results provided by search engines, especially when they do not have the skills or resources to access and manage this information intelligently. Furthermore, online users use keywords and very simple terms for their search, and they assume that search engines will understand their queries. The majority of online search users do not behave strategically in their search, they wait rather than the search engine can find the answer for them, regardless of their own strategy (Leeder et al. 2016).

Indeed, search engines have developed greatly since the advent of Big Data phenomenon. The effectiveness of the search for information, especially on the Web, would be particularly related to the expertise use of search engine system, including the knowledge of procedures and online research tools, but also the strategies to use in the search for information, and to quickly and accurately assess the content quality and the credibility of data and information returned. The significant growth of information in the Internet requires more effective search tools that can distinguish relevant information from hundreds or even thousands of raw data. However, the quality of the results provided by traditional search engines is not always appropriate, in particular when the user request becomes increasingly complex (Aoulad Abdelouarit et al. 2015). Figure 1 shows the model of online search engine in Big Data environment.

As shown in this figure, the user accesses the online search page via the web browser. He grabs the keywords and valid the search form. The search engine intercepts the user request and it start searching on the Internet data based on keywords entered. The collected data is referenced, indexed, and ordered before presenting them to the user on the results page. The parsed data includes all

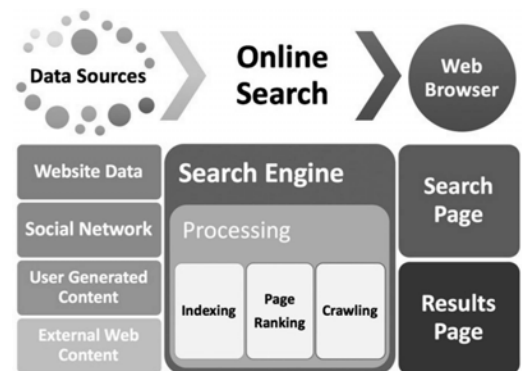


Figure 1. Online search model in Big Data environment.

internet data like: websites, social networking, data generated by users and other external data sources than the Web data. This great mass of heterogeneous data impacts negatively the user of the online search and makes the search process difficult to succeed. Thus, the use of a support system to search online information, based on the raw data of Big Data phenomenon presents major achievements. To do this, the need for a fictitious model to represent and process this type of data that are not necessarily text becomes essential.

2.2 *The problem of using massive and unstructured data generated from big data*

The massive and unstructured data on Internet is growing exponentially: If 20% of the data available on the web is structured data, the other 80% is unstructured (Goher et al. 2016). Unstructured data means that the elements within the data have no structure and they do not follow a specific or universal format manifesting the information that she wears. Unstructured data can include content that is similar or identical to corresponding structured data but not organized so that they are easily consumable, presentable, or be used by an application or user (Donneau-Golencer et al. 2016). Table 1 shows the different categories of unstructured data that circulate in the web with examples for each category.

As presented in this table, the information circulating on the Internet are manifested in several types (image, video, text, etc.) and comes from different sources (satellites, websites, social networks, etc.). While unstructured data do not take any form of organization, their information are so valuable that companies and researchers are constantly trying to find new ways to exploit them. Unstructured data can also contain digital information and factual details that may serve as a potential source of information (Nikhil et al. 2015).

3 PROCESSING MASSIVE AND UNSTRUCTURED DATA TO IMPROVE AND ORGANIZE ONLINE SEARCH RESULTS

3.1 *How to make massive and unstructured data consumable by users and applications?*

Unstructured data cannot be processed effectively because of their raw format. Thus, the information extraction techniques have been widely applied to extract important structured and manageable data from the original unstructured data. Indeed, unstructured data are processed by extraction solutions of structured data. The extracted data

Table 1. Categories of unstructured data generated from Big Data layer.

Data category	Examples
Satellite images	weather data, satellite surveillance imagery, etc.
Scientific data	seismic imagery, atmospheric data, and high energy physics.
Photographs/video	security, surveillance, and traffic video.
Radar or sonar Data	vehicular, meteorological, and oceanographic seismic profiles.
Text internal the company	documents, logs, survey results and e-to mails.
Social media data	YouTube, Facebook, Twitter, LinkedIn, Flickr, etc.
Mobile data	text messages and location information.
Website content	any site delivering unstructured content, like Wikis, YouTube, or Instagram.

provide summaries and sketches from the original unstructured data. There must be a loss of information after processing data reduction. However, in many applications, unstructured data is so large that a small summary can be precise enough to meet the needs of the analysis and processing of data (Chen et al. 2013). With the variety, velocity and volume of data circulating on the Web, it has become increasingly difficult to find patterns that lead to meaningful conclusions based on these data. Thus, the solution for processing massive and unstructured data may be performed in several steps:

Integration and cleansing: Data integration is the process of standardizing the data definitions and data structures of multiple data sources by using a common schema thereby providing a unified view of the data. Data cleansing is the process of detecting, correcting or removing incomplete, incorrect, inaccurate, irrelevant, out-of-date, corrupt, redundant, incorrectly formatted, duplicate, inconsistent, etc. records from a record set, table or database. Data cleansing is considered a major challenge to the Big Data era because of the increasing volume, velocity and variety of data in many applications.

Reduction: it consists in to reduce the countless amounts of data based on significant parts. It is the transformation masses of data or information, usually empirically or experimentally derived, into corrected, ordered, and simplified form, a kind of summarized reports. Big Data opens new opportunities and challenges to these techniques that have been well studied in the past and which include the approach of machine-learning that is a possible

way to improve traditional techniques for reducing data to process massive data.

Query and indexing: Indexing is sorting a number of records on multiple fields allowing searches and queries to be performed on it. In Big Data era, methods of indexing and searching with small structured data sets are no longer adequate. The tree structure is very popular in traditional indexing. While in the field of large volumes of data, this approach does not work well to provide simultaneous read and write operations without bottlenecks in the data structure. The need for a technical and effective method to improve data query and indexing from Big Data is essential.

Analysis and exploitation: Data analysis is the process of inspecting and modeling data with the goal of discovering useful information. Exploitation is to gain valuable and actionable insights from large and ever more complex data. However, considering heterogeneity and massiveness of data generated from big data, analysis and exploitation becomes very difficult. Typical relational database technologies meet a lot of difficulties when they are used to meet the challenges of deep analysis of massive data, because of their limited ability to expansion. Figure 2 shows the appearance of the already mentioned solution involving the processing of massive and unstructured data generated by the

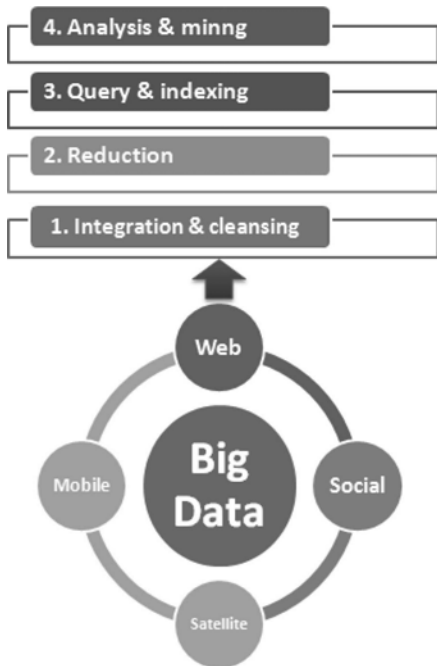


Figure 2. Processing solution for massive and unstructured data generated by Big Data layer.

Big Data layer so that they are easily consumable and presentable to users and applications.

As shown in this figure, the massive and unstructured data generated from Big Data layer must undergo treatment integration and cleaning at the beginning. Then, the reduction technique helps to avoid querying and indexing of large data retrieved from the Big Data layer. Finally, analysis and

Table 2. Processing tools for massive and unstructured data.

Solution	Description
Hadoop	Apache Hadoop is a framework for distributed storage and processing of large sets of structured and unstructured data. It is an open source project based on Google and uses the MapReduce algorithm for data processing.
MapReduce	It's the basis of the Hadoop technology and represents the combination of the two techniques Map and Reduce, where the Map() function is the master node that takes the input data and divides it into small parts to be distributed in different nodes. On each node a recursive operation is performed which leads to the tree structure to multiple levels. The result of treatment is returned to the master node. In the Reduce() function the master node collects the solutions from all the little nodes and merges them together to form the output result.
NoSQL	It allows manipulation of data without a prior defined pattern. These are semi-structured data or unstructured. This system avoids the schema definition and the loading of new data after cleaning. Its main features are: the horizontal extension of the system to perform simple operations distributed and data partitioning and replication on multiple servers, a more flexible concurrency model than traditional transactions, the indexing and distributed memory storage and easy changes in the data structure.
Machine-learning	It allows system to analyze hundreds of variables simultaneously, with their interconnections, to form patterns. It is well suited to complex problems involving multiple variables, it sticks very well with large and unstructured data, including images, text, audio, sensor data, etc. However, this approach is limited by its rules and does not provide options to well represent the processed data.

exploitation of information will focus on a small part of data, which will take less time processing and will provide more relevant information.

3.2 Towards a fictitious model for processing massive and unstructured data

Among the major challenges of the Big Data phenomenon, there is the treatment and representation of its massive and heterogeneous data. Since it is no longer a question of process data in standard format and use generic models to represent their content, the trend today is the use of data representations that promote rapid handling, storage and recovery of raw data, distributed and heterogeneous (Adiba et al. 2016). The real challenge is not to capture this masse of various data, but to analyze these data and exploit the most valuable pieces of information. Table 2 shows the most used techniques for the treatment of massive and unstructured data.

In addition to the comparison table just presented, we can say that NoSQL systems can provide too simple indexing strategies relative to RDBMS. They encourage the programming of queries using MapReduce model unlike declarative approach and optimized relational DBMS. In contrast, Hadoop can process large amounts of data directly, without defining a schema as for relational DBMS and it uses the MapReduce technique for that (Adiba et al. 2016). Thus, we can say that the most appropriate model for dealing with massive, heterogeneous and unstructured data resulting from online search, is implemented by the Hadoop-MapReduce technique. This technique allows to capture and store huge amounts of unstructured data in its native format.

4 TOWARDS AN APPROACH BASED ON HADOOP TO IMPROVE AND ORGANIZE MASIVES AND UNSTRUCTURED DATA

4.1 The integration of the online search with the hadoop technique

Today, the major challenges of online search engines manifest in understanding capacity, speed and accuracy in the collection and evaluation of search results expressed by the user. Almost of the search engines are based on the PageRank algorithm to assess the significance of the sites covered, this enhances greatly search engine accuracy, but the value of the rendered content does not always match the needs of the user. In addition, the actual search engines need to deal with huge amounts of data and complex calculations that emerge on Internet daily. The Hadoop technology alone does not provide a complete data searching system, but it is a distributed system programming tool that

essentially consists of two parts: The distributed model MapReduce and the distributed file system HDFS (Chen et al. 2014). Consequently, we can represent the previous figure 1 about the architecture of a complete system for online search, by replacing the conventional data processing layer of the search engine by the Hadoop layer for massive data processing. Figure 3 below shows the new model architecture implemented by the Hadoop-MapReduce technique to treat the massive and heterogeneous data and the interfacing of this model with the online research process.

This figure represents the typical architecture of a fully integrated online search system for interactive exploration of massive, heterogeneous data coming from several sources: websites clickstream, user generated content, content management system, external Web content, etc. This solution is based on the Hadoop Framework. Data sources of Big Data are stored in the Hadoop Distributed File System (HDFS). Using the MapReduce technique, these data will be processed, reduced and indexed to make them so well-suited for search, and therefore, provide a better platform for data mining than relational databases. It has to allow a natural language search using keywords and interactive navigation through its interface without additional training or advanced knowledge of programming. The main features should include scalable storage in HDFS, indexing batches via MapReduce to create scalable index for data stored in HDFS. This system should also allow the real-time indexing when collecting data and make it searchable when stored.

4.2 The use of a system based on hadoop for online search in big data environment

The Hadoop Framework is normally used to process massive and heterogeneous data. It can perform various operations such as data analysis,

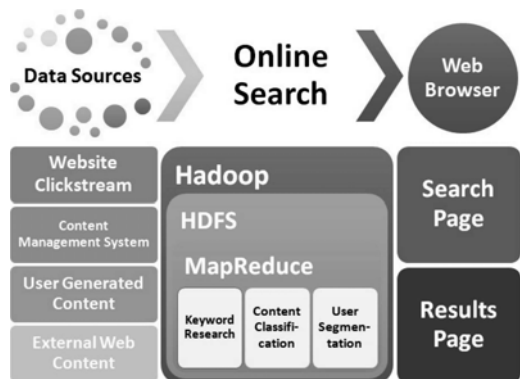


Figure 3. Integrating Hadoop-MapReduce with online search system.

analysis of results, etc. However, Hadoop can be combined with other techniques to offer an online search system to explore massive and disparate data coming from multiple internet sources including structured and unstructured data. Table 3 presents a set of solutions based on Hadoop technology and offering an online search system in an environment of massive and heterogeneous data.

According to this comparative table of solutions, we can say that to meet our need for improvement of online research in the Big Data environment, the best solution is to combine Hadoop technology for the storage and processing of massive and heterogeneous data, with a research framework as Solr, in addition to an indexing engine data like Lucene. Moreover, a specific development is needed in terms of this system presentation layer to meet the ergonomic needs and formatting of search results returned by the online search system. Figure 4 shows the architecture of the new solution with the

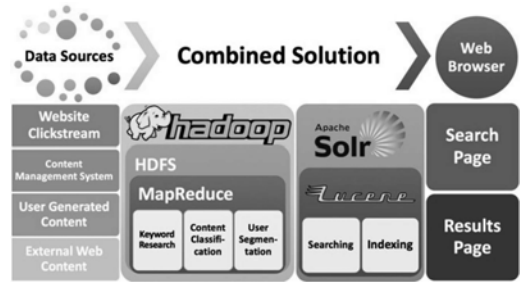


Figure 4. Integrating Hadoop-MapReduce with the Solr search framework.

Table 3. Online search solutions based on Hadoop.

Solution	Description
Solr	Is an open source search platform based on Apache Lucene project. It includes full text search. It uses the Lucene as search library for full-text indexing and search. It provides distributed indexing, replication and load-balanced querying, automated failover and recovery, centralized configuration.
Lucene	Is an open source Apache project that offers a text search engine library. It includes indexing, ranked searching, powerful query types like phrase queries, wildcard queries, proximity queries, range queries, fielded searching such as title, author, contents.
Elasticsearch	Is an open source, distributed, analytics, real-time search engine. It uses Lucene internally to build its distributed search and analytics capabilities. Elasticsearch clusters detect and remove failed nodes, and reorganize themselves to ensure that the data is safe and accessible.
Nutch	Is an open source web search engine based on Lucene for the search and index component. It is a highly extensible and scalable web crawler software project.
Cloudera	Is an open-source Apache Hadoop distribution, CDH (Cloudera Distribution Including Apache Hadoop), targets enterprise-class deployments of that technology.

use of Solr framework via its search interface, for searching data generated by the Big Data layer and processed by the Hadoop technology.

As presented in this figure, data coming from Big Data are intercepted by Hadoop layer that implements both a Distributed File System (HDFS) and an execution layer that supports the MapReduce programming model. Thus, data is loaded and transformed during the map phase, and then combined and saved during the reduce phase to write out Lucene indexes. The Lucene layer reads stored data from HDFS, and stores them using a Lucene Scheme, which in turn saves records as Lucene documents in an index. Once all of the files are indexed at Lucene layer, we can now perform queries against them. At Solr layer, we need to create a schema that matches the index that we are generating from Lucene layer.

4.3 Results and discussion

We can directly access Solr admin console by pointing our browser at: <http://<your-solr-server>:8983/solr/admin>. And from here we can run queries against Solr like the Figure 5 shows.

As shown in this figure, the response is in JSON format by default, and it found (in this example) 17 matches in 0ms for products that have the property `inStock=true`.

The combination of Hadoop and Solr allows to easily explore a lot of data, and then provide the results quickly via a flexible search interface. Solr supports multiple style queries, we can say that it replaces the NoSQL system for traditional databases, especially when the data size exceeds what is reasonable with a typical RDBMS. However, Solr has some limitations such as:

- Updating the index Lucene generates a new segment, which impacts performance;
- The replication feature is not yet supported in its Cloud (SolrCloud);
- Many SQL queries cannot be easily expressed with Solr queries.



Figure 5. Executing queries from the Solr admin console.

5 CONCLUSION AND FUTURE WORK

The growth of unstructured data is a particular challenge for Big Data in addition to the volume and diversity of data types beyond the capabilities of older technologies such as relational databases. Companies and researchers are constantly exploring the next generation of technologies for the analysis of such data. One of the most promising technologies is Hadoop and Apache MapReduce technology for managing massive and heterogeneous data. Thus, to improve online search results used in the massive, heterogeneous data environment, we must think to design a system based on Hadoop-MapReduce technique. As we have already mentioned, Apache offers several solutions in this context, which include the Solr system, Cloudera, Nutch, etc. In this paper, we propose the search Framework Solr combined with the Lucene indexing engine to implement the online search in the Big Data environment.

As perspective of this work, we intend to implement the scenario of the use of online search by integrating one of the solutions already mentioned. Then, in a second step, we will study the possibility of integration of our solution to improve teaching and scientific research for learners in e-learning environment, that was the subject of our previous study within the University of Abdelmalek Essaadi (Aoulad Abdelouarit et al. 2015).

REFERENCES

- Adiba, Michel, Castrejon-Castillo, Juan-Carlos, Oviedo & Javier Alfonso Espinosa et al. 2016. Big Data Management Challenges, Approaches, Tools and their limitations. *Networking for Big Data*.
- Aoulad Abdelouarit, Karim, Sbihi, Boubker, Aknin & Noura. 2015. Big-Learn: Towards a Tool Based on Big Data to Improve Research in an E-Learning Environment. *International Journal of Advanced Computer Science and Applications (IJACSA)* 6(10): 59–63.
- Chen, Jinchuan, Chen, Yueguo, DU & Xiaoyong et al. 2013. Big data challenge: a data management perspective. *Frontiers of Computer Science* 7(2): 157–164.
- Chen, Ning & Chai Xiangyang. 2014. Investigation on Hadoop-based Distributed Search Engine. *Journal of Software Engineering* 8 (3): 127–131.
- Donneau-Golencer, Thierry, Nitz, Boubker, Aknin & Kenneth C. 2016. Extracting and leveraging knowledge from unstructured data. *U.S. Patent No 9,245,010*.
- Gayathri, J. & Saraswathi, K. 2013. Extraction Of Data From Streaming Database. *International Journal of Computer Trends and Technology (IJCTT)* 4(10).
- Goher, S. ZerAfshan, Javed, Barkha, Bloodsworth & Peter. 2016. A Survey of Cloud-Based Services Leveraged by Big Data Applications. *Managing and Processing Big Data in Cloud Computing*: 121.
- Lakhani, Ajeet, Gupta, Ashish & Chandrasekaran, K. 2015. IntelliSearch: A search engine based on Big Data analytics integrated with crowdsourcing and category-based search. *Circuit, Power and Computing Technologies (ICCPCT), 2015 International Conference on. IEEE*: 1–6.
- Leeder, Chris, Shah & Chirag. 2016. Measuring the Effect of Virtual Librarian Intervention on Student Online Search. *The Journal of Academic Librarianship* 42(1): 2–7.
- Matei & Laura. 2014 Big Data Issues: Performance, Scalability, Availability. *Journal of Mobile, Embedded and Distributed Systems* 6(1):1–10.
- Nikhil, R., Tikoo, N., Kurle, S., Pisupati, H. S., & Prasad, G. R. 2015. A survey on text mining and sentiment analysis for unstructured web data. *Journal of Emerging Technologies and Innovative Research (JETIR)* 2(4).
- Sbihi, Boubker, El Kadiri, Kamal Eddine, Aknin & Noura. 2010. Towards a participatory E-learning 2.0 A new E-learning focused on learners and validation of the content. *arXiv preprint arXiv:1001.4738*.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

The efficacy of using electronic educational bag in developing functional expression skills among secondary school students at Bisha Governorate, Saudi Arabia

Mohamed Saeed Abdullah Al Mosaar

Education Ministry, Bisha Governorate, Bisha, Saudi Arabia

ABSTRACT: The objective of this article is to identify the efficacy of using electronic educational bag in developing functional expression skills among secondary school students at Bisha Governorate. The importance of this study is the compatibility with new trends in teaching, which focuses on using information technology in teaching. This study may be useful to those responsible for curriculum planning and development in designing Arabic language books for some modules in the form of educational bags, and also the outcome of this study may contribute to increase the students, teacher and also the school management attention to the expression subject. To achieve the objectives of this study; a list of functional expression skills was prepared and electronic educational bag was designed using several computer programs, and preparing achievement test, note card and ensure its validity and reliability. Experimental method has been used with quasi-experimental design includes two groups: Experimental group (which used electronic educational bag), Control group (which used the traditional method). The study sample consisted of 48 students; distributed equally on the study groups (Experimental and Control group) each group includes 24 students, the study results concluded the efficacy of electronic educational bag in developing functional expression skills among secondary school students at Bisha, the results also showed a statistically significant differences at the significance level of ($\alpha \geq 0.05$) between the mean scores of the Experimental and Control group in the post measurement of educational achievement and functional expression skills for the Experimental group which learned with electronic educational bags. The most important recommendations of this study is to activate the electronic educational bag use in teaching the Arabic language and various curriculums, and conduct similar studies on the effect of using electronic educational bags on educational achievement and development functional expression skills among secondary school students and other variables.

1 INTRODUCTION

Educational bags are considered a strategy based on the principle of self-learning also, and focus on the presence of various means, alternatives and methods of teaching for a learner who could achieve his desired educational goals through practicing such means as a learner follows it according his educational speed and self-step, also give slow learners more time in order to achieve their desired educational goals without feeling a failure as such means do not based on comparing a learner educational level with any other learner, the only standard here is to reach the required level of proficiency [1].

Teachers could construct electronic bags in all subjects at any level and any educational stage. Electronic bag can be seen as ordered record and model or a sample of student work in the curriculum of various subjects in order to give an important dimension which is the student learning in all

aspects, measures the student skills and knowledge in many ways. Such bags include curriculum, calendar and classroom teaching in an integrated manner. Through using electronic bag; students can build and develop their understanding of the good work and classrooms could be based on student more than teacher [2].

Expression subject is one of the most important language skills and goals in the branches of language that educational process aims to achieve, but the other branches of language helping to reach this end. As rules are ways to preserve the tongue and pen from error in the expression, dictation and handwriting are means to draw words and sentences in a right and clear way for expression. Reading provides the reader with the language grammar, knowledge and culture that enriches the ability to express. All these branches at the end are means to improve the quality of expression in both types of oral and written at their different purposes whether functional or creative. The foregoing showed the

importance of functional expression for secondary school students and the importance of using educational bags in providing information or skill so that, this study aims to identify the efficacy of electronic educational bags in teaching functional expression subject and developing expression skills for secondary school students in the second year. In this paper; the study problem, methodology, sample and importance will be shown as well as the literature review, results in numbers, and finally results, recommendations and proposals.

2 PROBLEM FORMULATION

Through the researcher considering for the students levels of expression in educational classrooms as the researcher works in the field of education, in particular the researcher specialization in the same field of study which is the Arabic language, and after reviewing the students' records in the previous classrooms; the researcher found the student's low levels and defaults performing skills related to expression, in general, and functional expression in particular.

Expression is one and the most important goals of teaching Arabic language. All branches of Arabic language aim to make the learner a good expressive by having the ability to transmit his ideas and feelings to others. The scientific and cognitive progress in all aspects require the individual to have the ability that enables him to deal with this progress, especially in the functional and practical aspects which require specialists and educators in turn to give their attention to the learner, developing his functional expression skills, search for new strategies and methods of technology that helps in developing such skills so that; the learner could achieve the desired objectives from learning expression in order to keep pace with this progress.

The study problem summarized in answering on the following key questions; the first question:

"How using electronic educational bag effect in developing academic achievement among secondary school students at Bisha Governorate?"

The second question: "How using electronic educational bag effect in developing functional expression skills among secondary school students at Bisha Governorate?"

3 THE STUDY GOALS

This article sought to identify functional expression skills among secondary school students at Bisha and identify the efficacy of electronic educational bags in providing functional expression skills for secondary school students at Bisha, as well as identify the efficacy of electronic educational bags

in increasing the academic achievement of secondary school students at Bisha.

4 IMPORTANCE OF THE STUDY

The importance of this paper is the compatibility with new trends in teaching, which focuses on using information technology in teaching. This study may be useful to those responsible for curriculum planning and development in designing Arabic language books for some modules in the form of educational bags, and also the outcome of this study may contribute to increase the students, teacher and also the school management attention to expression subject.

5 METHODOLOGY OF SOLUTION

The researcher used the experimental method in this study in order to detect the efficacy of electronic educational bag in developing functional expression skills among secondary school students at Bisha. This study based on quasi-experimental design with two groups. Figure (1) illustrates the same:

The researcher used the achievement test to measure the cognitive aspect of skills and to ensure the equality and homogeneity of both groups, and also used note card to measure the functional expression skills of secondary school students at Bisha, in addition to the statistical program SPSS. There are two hypotheses in this study. The first "There are no statistically significant differences at the level of ($\alpha \geq 0.05$) between the scores mean of the Experimental and Control groups in the post measurement of the academic achievement of secondary school students at Bisha". The second "There is no statistically significant differences at the level of ($\alpha \geq 0.05$) between the scores mean of the Experimental and Control groups in the post measurement of functional expression skills of secondary school students at Bisha".

6 IMPORTANT SIMILAR WORKS: SURVEY STUDY

The first study [3] entitled "The efficacy of teaching using educational bags as a style of individual learning in the achievement of fourth grade pupils", which aimed to identify the efficacy of teaching using educational bags as a style of individual learning styles in the achievement of fourth grade pupils in Jordan (Experimental Group), and comparing the same with the academic achievement of the control group who learning in traditional method. The study sample consists of (58) students: the experimental group includes

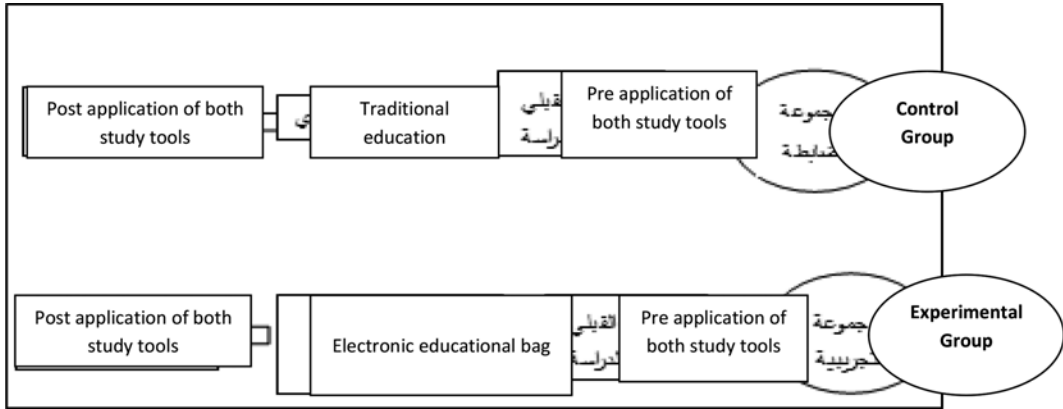


Figure 1. Quasi-experimental design of the study.

(29) students which learned using educational bags, and the control group includes (29) students which learned using the traditional method in the second term 2009/2010. In order to test the efficacy of the bag; pre-test and post-test were prepared to be applied on the two groups. Both researchers noted a statistical differences at the level ($\alpha = 0.05$) between the two groups mean for the experimental group which studied the subject through educational bags. In the light of the study results, both researchers recommended with the importance of designing some units in the books of social and national education for students in the primary school in the form of educational bags to be suitable with schools where such bags being applied.

This study is associated with the current study as it shows us the efficacy of educational bags as an effective educational style which enhances the impact and efficacy of educational bags.

The work [4] entitled “Learning Efficacy of Simultaneous Audio and on-screen Text in Online”, which aimed to identify the efficacy of learning texts displayed in two ways, the first way: using educational bags includes computer slides to display on-screen text, and the second way: displaying text through multimedia bags. The study was conducted on a sample of students in an area in Malaysia and pre-test and post-test were applied on the two groups so that, the importance of integrating the computational educational bags and regular multimedia educational bags within teaching process, taught through experience, was cleared. After applying the achievement test on the two groups, results and recommendations showed the superiority of the experimental group which learned through educational bags more than female students in the control group which learned through traditional method. The study recommended the need for training courses and

seminars for teachers in order to identify features of educational bag and apply it in teaching.

This study is associated with the current study as it shows us the superiority of educational bags in teaching through the experimental research which confirms the quality of the study direction.

This study [5] entitled “The effect of using educational bag on the achievement and retention of fourth grade female pupils for the geographical terms unit”, which aimed to identify the effect of using educational bag on the achievement and retention of fourth grade female pupils for the geographical terms unit. The researcher used the experimental method on a sample consisted of all fourth grade female pupils at “Al-Thalathon Elementary School for girls” in Medina. The researcher used an educational bag designed by her as well as pre-test, post-test and retention test also prepared by her. To find out the effect of using educational bag and verify the research hypotheses through tests conducted by the researcher; the results showed the statistically significant superiority of the experimental group more than the control group in the achievement and retention. The most important recommendations: holding training courses for social subjects’ teachers and the other subjects on designing and using educational bags and provide schools with necessary devices and materials for educational bags production.

This study is associated with the current study as it shows the quality of educational bags for the female pupils’ achievement of terms which constitute the basic of expression construction.

On other hand, the work [6] entitled “The effect of designing electronic bag on academic achievement and the parents and students satisfaction degree about it in science subject for primary school pupils”, which aimed to identify the effect of designing and using electronic bag in

science subject for primary school on the pupils achievement and the parents and students satisfaction degree about it. The study sample consists of (49) pupils in two schools “Al-Muthana & Ahmed Al-Khamis Primary School for boys” at Farwaniya Educational District in Kuwait in the first term 2009/2010. The experimental group consisted of 24 pupils in Al-Muthana Primary School and the control group consisted of 25 pupils in Ahmed Al-Khamis Primary School. The research results detected statistically significant differences in the achievement for the experimental group which studied using the electronic bag and a statistically significant high degree of satisfaction among students in the experimental group; but found low degree of satisfaction among parents regarding the electronic bag. This study is associated with the current study as it shows us the degree of satisfaction among students and their acceptability regarding the electronic bag as an educational style.

Finally the work [7] entitled “User-Education in a Flexible Learning Environment—An Opportunity to Stay Relevant in the 21st Century”, which aimed to identify the efficacy of using traditional educational bags with students in high schools in South Africa and the role of educators in developing and operating modern educational bags. The research was conducted on three groups of students; the first group used traditional styles of educational bags, the second group used improved and enhanced teaching methods on historical information in the 21st century and the third group provided with information through multimedia educational bags. The study results revealed that the third group had a high percentage of information retention. The study recommended the need to develop and modify educational bags through active learning and multimedia. This study is associated with the current study as it shows us the efficacy of educational bags as an effective educational style which enhances the effect and efficacy of educational bags.

7 NUMERICAL RESULTS

The first hypothesis: “Are there any statistically significant differences at the level of ($\alpha \geq 0.05$)

between the scores mean of the Experimental and Control groups’ students in the post measurement of academic achievement?”

To identify the statistically significant differences between the scores mean of the Experimental and Control groups’ students by performing the post measurement in the academic achievement test among secondary school students at Bisha; the researcher used T test for independent samples on the study samples (the experimental and control) after applying the electronic bag. The following table illustrates the results:

From the above Table 1; there are statistically significant differences between scores of the control and the experimental group in the post measurement of the academic achievement at the significant level of 0.01 or less. It is clear that using the proposed electronic educational bag by the researcher had an efficacy and positive impact on increasing the academic achievement among secondary school students at Bisha so that, the first hypothesis was wrong and the alternative hypothesis was accepted which supposes that: There is statistically significant differences between the scores mean of the Experimental and Control groups in the post measurement of academic achievement among secondary school students at Bisha for the experimental group.

The second hypothesis supposes that: “Are there any statistically significant differences at the level of ($\alpha \geq 0.05$) between the scores mean of the Experimental and Control groups’ students in the post measurement of functional expression skills?”

To identify the statistically significant differences between the scores mean of the Experimental and Control groups’ students by performing the post measurement of functional expression skills among secondary school students at Bisha; the researcher used T test for independent samples on the study samples (the experimental and control) after applying the electronic bag in the cognitive aspect. The following table illustrates the results:

From the above Table 2; there are statistically significant differences between scores of the control and the experimental group in the post measurement regarding skill aspect in the functional expression skills at the significant level of 0.01 and less.

It is clear that using the proposed electronic educational bag by the researcher had an efficacy and

Table 1. T test results for independent samples between the scores mean of the Experimental and Control groups according to the post measurement of the academic achievement.

Study variable	Group	No.	Mean	Standard deviation	T value	Degrees of freedom	Significance level
Post measurement of achievement	Experimental	24	13.96	3.42	-2.789	46	0.008**
	Control	24	16.33	2.39			

**Significant differences at the level of 0.01 or less.

Table 2. T test results for independent samples between the scores mean of the Experimental and Control groups according to the post measurement of the functional expression skills.

Study axis	Sample	No.	Mean	Standard deviation	T value	Degrees of freedom	Significance level
Skill aspect in the functional expression skills	Experimental group	24	27.38	7.04	-6.33	46	0.00** Significance
	Control group	24	39.00	5.60			

**Significant differences at the level of 0.01 and less.

positive impact on increasing the skill aspect in the functional expression skills among secondary school students at Bisha so that, the second hypothesis was wrong and the alternative hypothesis was accepted which supposes that: There is statistically significant differences between the scores mean of the Experimental and Control groups in the post measurement of functional expression skills among secondary school students at Bisha for the experimental group students.

8 CONCLUSION AND RECOMMENDATIONS

This study found the efficacy of electronic educational bag in increasing the academic achievement among secondary school students at Bisha, and also there are statistically significant differences between scores mean of the control and experimental group in the post measurement regarding the functional expression skills among secondary school students at Bisha by performing the academic achievement test for the experimental group. The electronic educational bag is also affective in developing the students' skills in the functional expression, and finally, there are statistically significant differences between scores mean of the control and experimental group in the post measurement regarding the skill aspect in functional expression skills for the experimental group.

This work recommended with producing electronic educational bags under the supervision of the Ministry of Education based on individual characteristics in order to reach an adaptive learning for each learner to meet their needs and achieve its objectives. This study also recommended with training learners on acquiring educational programs production skills using electronic educational bags because of its importance in developing both of cognitive and performative aspect of the curriculums and subjects.

This study proposed to conduct a study about "The difficulties faced by learners when training on the production of electronic educational bags",

study about "The efficacy of using electronic educational bags in the academic achievement and developing functional expression skills for students in Primary and Elementary school", and also conducting study about "The effect of using a software based on the educational bag in the academic achievement at various educational stages", in addition to a comparative study between traditional educational bag and electronic educational bag and identify the effect of each bag on the academic achievement and acquiring skills, and finally conducting a similar study to the current study in developing other skills for students.

REFERENCES

1. Saraya, Adel (2007): "Individualized Education Technology and Innovation Development" (Ed. 1), Amman: Dar Wael for Publishing and Distribution.
2. Zaiton, Kamal Abdel-Hamid, (2004): "Education Technology in the Age of Information and Communication", World of books: Second Edition.
3. Al Titi, Maram Amer, Hawamdah, Nada Mohamed (2010): "The efficacy of teaching using educational bags as a style of individual learning styles in the achievement of fourth grade pupils", unpublished Master Thesis, The Hashemite University, Jordan.
4. Debusse, Justin C. W; Hede, Andrew; Lowley, Merdedith (2009): "Learning Efficacy of Simultaneous Audio and on-screen Text in Online".
5. Masoudi, Amal (2003): "The effect of using educational bag on the achievement and retention of fourth grade female pupils for the geographical terms unit", Unpublished Master Thesis, Dean of Graduate Studies, King Abdul-Aziz University, Saudi Arabia.
6. Anzi, Mishal Mohamed (2009): "The effect of designing electronic bag on academic achievement and the parents and students satisfaction degree about it in science subject for primary school pupils", Research in the field of e-learning, Arabian Gulf University.
7. Van, Vyrebm A. J; Henning, J. C. (1998): "User-Education in a Flexible Learning Environment—An Opportunity to Stay Relevant in the 21st Century", International Association of Technological University Libraries (IATUL) Conference (Pretoria, South Africa, June 1–5, 1998), Volume 18; see IR 057 503.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

A proposed real-time EMG intelligent robot control algorithm based on swarm intelligence and neural networks

Bassant Mohamed ElBagoury

*DeanShip of Preparatory Year, Taif University, Taif, Kingdom of Saudi Arabia
Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt*

Jehad Al-Amri

College of Computers and Information Technology, Taif University, Taif, Kingdom of Saudi Arabia

Mohamed Roushdy

Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt

ABSTRACT: Controlling autonomous, rehabilitation robots for patients in a dynamic, continuous, and real-time environment is a complex task, especially motion control and patient EMG muscle signal feedback problems. This paper presents two integrated parts, which are a novel hybrid EMG classifier—for rehabilitation robotics control. The hybrid classifier includes kalman filter, Parallel Fish Swarm Algorithm (PFSA) and support vector Neural controller for rehabilitation robot. It depends on real-time EMG sensors data and wireless sensor board implemented for the telemedical controller. Our experiments are conducted in Simulink/matlab. The integration of kalman filter, fish swarm optimization for support vector machines neural network module enhances real-time EMG classification accuracy that reaches an average of 92.3%. The algorithm modules, results and testing are presented in this paper.

Keywords: Fish Swarm Algorithm, kalman filter, Support Vector Machines, EMG Shimmer Sensor, Robot Control

1 INTRODUCTION

Stroke and cardio vascular diseases have a high incidence in many countries. Beside the early detection of high-risk persons, their monitoring and the detection critical, death trap events, their effective emergency management the rehabilitation process is difficult and cost intensive.

Although stroke is a disease of the brain, it can affect the entire body. A common disability that results from stroke is complete paralysis on one side of the body, called hemiplegic. A related disability that is not as debilitating as paralysis is one-sided weakness or hemi paresis. Many stroke patients experience pain in legs and hands. Therefore, patients' rehabilitation treatment is very important for long time recovery and overall patient health management.

The rehabilitation points towards the intense and repetitive movement assisted therapy that has shown significant beneficial impact on a large segment of the patients. The availability of such training techniques, however, are limited by

- the amount of costly therapist's time they involve,
- the ability of the therapist to provide controlled,
- quantifiable and repeatable assistance.

These limitations are quite important. Rehabilitation robotics systems are a very important problem, especially in the therapeutic domain of stroke patients. This is due to:

- The complexities of patients' treatments procedures such as physiotherapy.
- Since Electromyography (EMG) detects muscle response during different actions, it gives useful identification of the symptoms' causes. Such disorders that can be identified by EMG are neuromuscular diseases, Nerve injury, and Muscle degeneration. The dealing with Electromyography (EMG) signals provides significant source of information for identification of neuromuscular disorders.
- A robot-assisted rehabilitation can provide quantifiable and repeatable assistance that ensure consistency during the rehabilitation and

- A robot-assisted rehabilitation is likely to be cost-efficient.

The aim of this paper is to extend our previous development of an Intelligent Telemedical Robotic rehabilitation platform [2] and enhance the intelligent control engine with real-time EMG signal feedback and artificial intelligence technologies.

2 RELATED WORK

Many researches were made in the field of EMG signal classification using different techniques. In [4] one against all method multi-class SVM with Gaussian kernel function was implemented to identify six degree of freedom. The overall rate of correct class testing was 97%.

In [4] Support Vector Machines (SVM) was employed to extract classes of different force intensity from the EMG signals. The average accuracy reached about 96%.

In [5] four Electromyography (EMG) sensors were used. They placed at the thigh and two Force Sensing Resistors (FSR) placed below the heel and the toe. Support vector machine was used to detect muscular activity changes. This system has reached accuracies of roughly 67% for an amputee and of 75% for a non-amputee individual.

Research has shown that [6] [7] [8] the kernel function parameters affects the SVM affect its classification ability. When the values of the Kernel function parameters factor are appropriate, the classification of SVM will enhance significantly.

Jing bang et al. [9] apply Parallel Fish Swarm algorithm for SVM optimization. In this paper, we utilize their method and moreover, we enhance it by new approach of kalman filter EMG signal estimate.

3 DESCRIPTION OF THE PROPOSED REAL-TIME EMG ROBOT CONTROLLER

Fig. 1. Shows the main components of the proposed intelligent hybrid rehabilitation robot controller based on a Telemedical platform for a portable rehabilitation robot monitor system. The Telemedical platform allows to manage the monitoring of high-risk patients of, detect critical events and control the rehabilitation process using wireless sensors and robots. The proposed controller consists of:

1. EEG, ECG, EMG wireless sensors signal acquisition and feature extraction.
2. Discrete Kalman Filter module for Estimating EMG signal motions.
3. Hybrid Adaptive behavior Robot controller developed with Fish Swarm Algorithm and SVM module for Robot Control.
4. A NAO Humanoid Robot Therapist for testing interaction with the patient motions.

Fig. 1. Shows the full system units, wireless telemedicine unit, signal processing and feature extraction units, robot controller unit system.

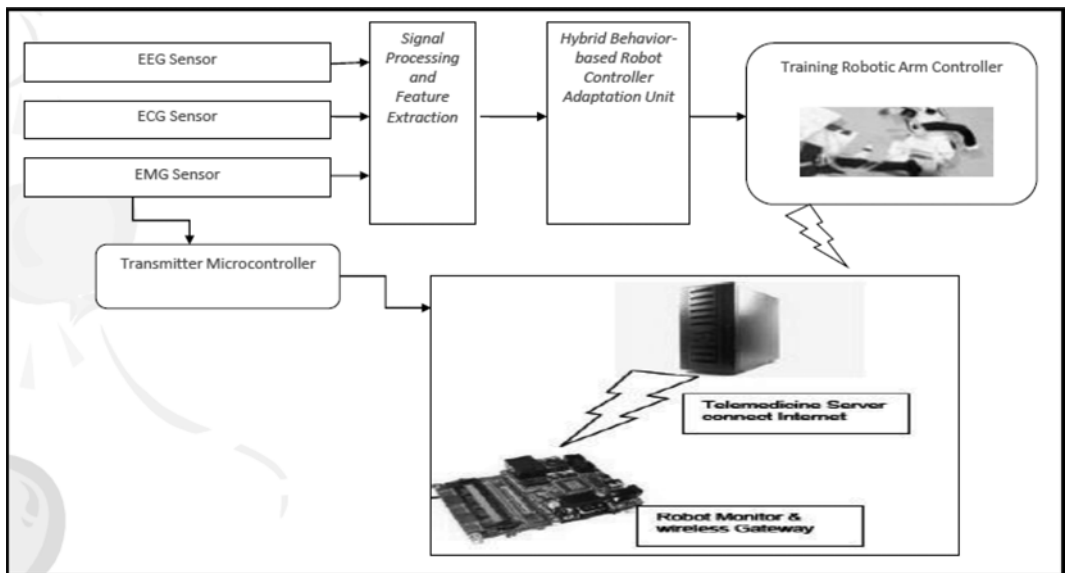


Figure 1. Intelligent telemedicine rehabilitation robotic architecture.

In this paper, we focus only on EMG signal processing, Kalman filter signal estimation and the novel adaptive robot control developed with Fish Swarm Algorithm and SVM optimization. These are described in details in coming sections.

3.1 EMG sensors and wireless sensor network

The Wireless Sensor Network was build up by usage of commercial available Shimmer-Sensors [10]. A sensor consists of a motherboard with a Class 2 Bluetooth radio module and can be added by a daughterboard for various applications: Electromyogram (EMG) and Electrocardiogram (ECG). The EMG sensor is combined with a motion detector consisting of a gyroscope and an accelerometer for angular and linear velocity. The EMG provides pre-amplification of the signal with a sampling rate of 512 or 1024 Hz. The signal range lies between 4 mV and +4 mV. The EMG sensors were placed on eight muscles: M. biceps femoris, M. vastus lateralis, M. tibialis anterior and M. soleus (each on both sides). The skin preparation as well as the placement of the Ag/AgCl electrodes follows the recommendations of the SENIAM Project. Only the gyroscope data were used and contains 3-axis data of angel velocity (x, y, z or roll, yaw and pitch). The sensor node data were collected by the MultisyncSoftware by Shimmer with a sampling rate of 512 Hz. This means for the EMG signals the signal acquisition covered frequencies up to max. 256 Hz.

3.2 EMG signal acquisition

The EMG sensors were placed on eight muscles. M. biceps, femoris, M. vastuslateralis and M. titbisanteriors. The sensor node data were collected by teMultisync-software by shimmer with a sampling rate of 512 Hz. This means for EMG signals the signal acquisition covered frequencies up to max. Fig. 2. Shows sample shoot for placement of sensor.

3.3 EMG feature extractions

The analytical process of EMG covers three different aspects: time-analysis, frequency-analysis and dynamical analysis. Task of time analysis is the detection of activity and rest state, the duration and characteristics of these states (amplitudes, mean of activity, standard deviation). The meanwhile standard procedure in EMG-signal processing is [7], [8]:

- rectification of the signal;
- decorrelation EMG from DC;
- creation of envelope curve of EMG.

We used a low pass, equiripple FIR-filter at 10 Hz (different frequencies are reported in literature [13]–[14]). The envelope curve increases the possibility to differentiate between activity and rest phases and can be used to detect these phases automatically by application of a threshold. In this study a separate rest EMG was measured as basis EMG. The threshold was calculated by the mean of rectified signal + 3 x standard deviation.

A. Time domain analysis

The time analysis evaluates the behavior of the emg signal over time: e.g. frequency of activity phases, their duration and intensity. The following parameter can be calculated easily [15], [16]:

- Amplitude of each activity state and mean of a set. Of activity states including their standard deviation.
- Duration of activity states.
- Area under the curve in activity states.
- Velocity of onset and onset of activity state.
- Comparison between different types of efforts and Their activity states in contrast to the rest state and between them.
- Time related pattern of activity states in case of

B. The frequency analysis

It can be used to detect various aspects of the signal. Intensity of activation and fatigue of muscle correlate with pattern of frequency and their changes. Two main frequency domain parameter in EMG signals are [17], [18], [19], [20]:

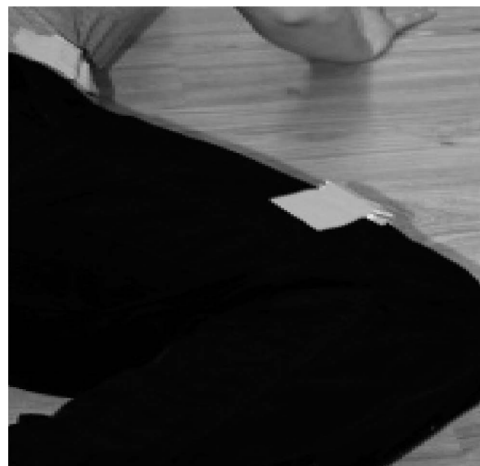


Figure 2. Sample shoot of shimmer sensor placement on leg muscles.

1. Mean Frequency (MNF):

$$MNF = \frac{\sum_{i=1}^N f_i P_i}{\sum_{i=1}^N P_i} \quad (1)$$

2. Median Frequency (MDF):

$$MDF = \frac{1}{2} \sum_{i=1}^N P_i \quad (2)$$

Other parameter can be derived from the frequency domain.

3. Total Power the aggregation of power spectrum (TTP) corresponds to the zero statistical moment [21], [22]

$$TTP = \sum_{i=1}^N P_i \quad (3)$$

4. Mean Average Power of EMG signal (MNP) [23], [24]

$$MNP = \frac{\sum_{i=1}^N P_i}{N} \quad (4)$$

5. Peak Frequency the frequency with the maximum Of power [25], [26]

- Spectral Moments of power spectrum [27], [28]
 - SM0 corresponds to TTP
 - SM1, SM2, SM3

$$SM1 = \sum_{i=1}^N f_i P_i \quad (5)$$

$$SM2 = \sum_{i=1}^N f_i^2 P_i \quad (6)$$

$$SM3 = \sum_{i=1}^N f_i^3 P_i \quad (7)$$

Beside the frequency analysis based on Fourier Transformation, the wavelet transformation can be quite useful in noise reduction, detecting very fast as well as very slow changes and self-similarity [29], [30].

3.4 Dynamic system analysis

The a linear system model. However, the effects of the neurological signal transduction at the neuromuscular junction and the recruitment of muscle fibers maybe not linear. The Recurrence Quantification Analysis (RQA) is a nonlinear method to analysis surface EMG signals and was used in

many other biosignal analysis domains such as electrocardiogram and encephalogram analysis. The result of RQU is a visualization of a square matrix representing system states over the time [31]. In a first step, the mono-channel EMG signal is transformed into the phase space by means of time delay procedure. The signal is shifted by a number of samples. The calculation of the shifting number can be done. by first zero of autocorrelation. Same states in phase space are visualized in the recurrence plot, can be quantified by histograms and expressed as:

Recurrence Rate:

$$RR = \frac{1}{N} \sum_{i,j=1}^{N^2} R_{i,j} \quad (8)$$

Determinism diagonal lines is the recurrence plots express a deterministic behavior and is calculated as:

$$DET = \frac{\sum_{I=I_{\min}}^N IP(I)}{\sum_{I=I,j}^N R_{i,j}} \quad (9)$$

The surface EMG is a summation of many single Muscle Unit Action Potentials (MUAP). MUAPS are repeating events for each muscle fiber. It is expected that more MUAPS are activated [32].

4 PROPOSED INTELLIGENT REAL-TIME EMG INTELLIGENT REHABILITATION ROBOTIC CONTROL

EMG signal classification plays the most crucial part in rehabilitation robotics systems [33]. It means analyzing and predicting muscle signals from patients with stroke to derive attended motion and control robot. In this section, a new hybrid Artificial Intelligence [34] algorithm has been developed for rehabilitation control. It is a modification of our SVM classifier [35] as it now includes Kalman Filter for EMG signal estimation and swarm intelligence for SVM optimization. The main proposed algorithm is shown in Fig. 3.

4.1 EMG signal classification

Various classification techniques have been proposed by many researchers. SVM [12] is a powerful learning method which aims to find the best the best hyper plane that can separate data perfectly into its two classes. Multi-classification was recently achieved by combining multiple SVMs.

Step A: 1. Apply Discrete Kalman Filter for knee Joint Angel Estimation.
Step B: 1. Input Real-Time EMG channels values $i=1..8$ to two SVM machines, one for classifying between kneeling pose 1, kneeling pose 2 and the second for classifying between kneeling pose 3 and kneeling pose 4.
 2. Input Extracted features values also.
 3. Apply Parallel Fish Swarm Algorithm PFSA to train SVM1
 4. Apply PFSA to train SVM2
 5. Fix SVM1 model and Fix SVM2 model
 6. Test Accuracy and repeat till best solution found
 7. Fix Best Architecture of SVM based on PSO optimization.
Step C: 1. Input EKF as estimated motion command to PFSA-SVM
 2. Output best estimation for Robot Control motion.

Figure 3. Proposed algorithm steps.

There are two schemes of SVM multi-classifier: (a) One Against All which classify each class against remaining classes and (b) One Against One which classify between each two classes. In our research, One Against one SVM classifier with Gaussian Radial Basis kernel Function (RBF) and sigma equal 1 was used in identification of EMG motions aggressive against ed SVM can be represented as

$$K(x_a, x_b) = \exp\left(-\frac{\|x_a - x_b\|}{2\sigma^2}\right) \quad (10)$$

4.2 Parallel fish swarm algorithm-SVM optimization—

Artificial Fish Swarm Algorithm is a kind of optimization strategy [ref.]. Parallel AFS is a modified one for fast optimization introduced by Jiang Bai et al [ref.] for SVM optimization for speech recognition system. SVM based on PFSA [ref] optimizes two important hyperparameters C and γ [34]. The hyperparameter C determines the trade-off between the model complexity and the degree to which deviations larger than ϵ are tolerated. A poor choice of C will lead to an imbalance between model complexity minimization and empirical risk minimization. The hyper-parameter ϵ controls the width of the ϵ -insensitive zone, and its value affects the number of SVs used to construct the classification function. If ϵ is set too large, this would result in too few SVs selected and lead to unacceptable “flat” classification estimates [10].

The specific steps of looking for the optimal solutions of parameters C and γ as adopted by Jiang Bai et al. [ref] are shown in Figure 4.

4.3 Kalman filter for knee joint angle estimation based on EMG sensor signal

Kalman Filter (KF) is widely used in studies of dynamic systems, analysis, estimation, prediction, processing and control[ref]. Kalman filter is an optimal solution for the discrete data linear filtering problem. KF is a set of mathematical equations which provide an efficient computational solution to sequential systems. The filter is very powerful in several aspects: It supports estimation of past, present, and future states (prediction), and it can do so even when the precise nature of the modeled system is unknown. The filter is derived by finding the estimator for a linear system, however, the real system is non-linear, and Linearization using the approximation technique has been used to handle the non-linear system. This extension of the non-linear system is called the Extended Kalman Filter (EKF) [1]. EKF have been extensively used in many applications where non-linear dynamics are prevalent. There are many instances where EKFs [1] have been used in different Robot controls [2, 3, 4]. In our previous work [2, 4], we have introduced complete Intelligent mission sensor model for rehabilitation robot control. It includes EMG and Accelerometer and other sensors. In this paper, we apply only two methods of EKF for knee poses

Step1: initialization. Generate an initial population which determine the scope of SVM parameter vector array (C, γ) .
Step 2: Train SVM with initial dataset
Step 3: Then the trained model is used to test the testing data set. The recognition rate of the overall testing samples is converted to the fitness value of the model.
Step 4: according to the fitness value, the next generation of parameter array (C, γ) population for optimization of the better (C, γ) is searched by PFSA[ref]
Step5: using the parameters of the offspring (C, γ) population, re-training and testing the SVM, calculates the corresponding fitness value. If it satisfies with the termination condition for training in PAFSA, it goes to
Step6: otherwise, it returns to Step3 to continue the operation.
Step7: ending training. Now, the parameters (C, γ) gotten is the final model parameters.

Figure 4. Parallel fish swarm algorithm adapted from [34].

based on EMG and Accelerometer sensors without including robot full kinematics.

Modified EKF method for knee joint angel

We modified the method proposed by F. Widjaja et al. [12], the joint angel $\theta(k)$ can be assumed to follow sinusoidal wave and we add MNF, mean frequency defined in eq(1) as follows:

$$\theta(k) = A_r \sin(2\pi \text{MNF}_t \text{K T}_s) \quad (11)$$

And angular velocity as $\dot{\theta}(k)$ as its derivative, also MNF has been used instead of signal frequency for averaging loss of lower frequencies as follows:

$$\dot{\theta}(k) = A_r 2\pi \text{MNF}_t \text{T}_s \sin(2\pi \text{F}_t \text{K T}_s) \quad (12)$$

where:

A_r is Amplitude of each activity state of sensor signal.

Therefore as derived by F. Widjaja et al. [12] the state model of the system at $(t+1)$ is:

$$\begin{aligned} X(t+1) &= \begin{bmatrix} \theta(k+1) \\ \dot{\theta}(k+1) \end{bmatrix} \\ &= \begin{bmatrix} \cos(2\pi \text{F}_t \text{K T}_s) & \sin(2\pi \text{F}_t \text{K T}_s) \\ -2\pi \text{F}_t \text{K T}_s \sin(2\pi \text{F}_t \text{K T}_s) & \cos(2\pi \text{F}_t \text{K T}_s) \end{bmatrix} \\ &= X(k) + V(k) \end{aligned} \quad (13)$$

Where: $V(K)$ is zero mean white Gaussian noise. And as it is defined as well in F. Widjaja et al. [12], the measurement model depends on EMG and ACC data.

5 DISCUSSION AND RESULTS

EMG data analysis allows the identification of activity and rest states, their properties (duration, amplitudes, area under the curve etc.) and their dynamical behavior. In the course of experiment with different tasks of kneeling motions the activity states have different durations and amplitudes

indicating different knee poses. Especially the phase of knee flexion increased the frequency and duration of activity states and changes of MNF can indicate an increase of knee poses. However integrating kalman filter for knee poses shows an increasing classification accuracy of the classifier as shown in Fig. 5. Nevertheless, these observed changes still requires full robot kinematics equations to be included for complete gait analysis. However, the results of the proposed hybrid classifier increased significantly. It seems that the synchronization of parameters enhances performance of robot controller.

Sensor readings

The kneeling motions are divided into four groups kneeling start pose, kneeling flexion, kneeling extension and kneeling normal pose. Each group consists of 1400 reading from 8 channels.

Cross-validation test

For experiments, cross-validation test uses 500 records are used for training the model and 500 for testing them. The results shown in Table 1 indicate the performance of applying different methods for each group.

Single joint single DOF

The flexion-extension movement as shown in Fig. 8, of the knee joint was selected in this experiment. This action is suitable for stroke patients in early rehabilitation training since it is easily be implemented by a robot model and simple enough for patients. The results of three methods SVM model and SVM with Swarm Intelligence and Kalman Filter with SVM and Swarm Intelligence neural network models are shown in Figure 5. As shown, in First Experiments the SVM architecture is basic one Against-one and it is trained on 9 features including MNF, MDF, TTP MNP, RR, DET, SM1, SM2, SM3 (defined before in section 3.3). In the second experiments, the SVM model is a hybrid with optimized parameters of Parallel Fist Swarm Intelligence. In third experiments, the SVM along with Swarm Intelligence is trained on one

Table 1. Shows a sample of EMG sensors readings.

Action	Ch1	Ch2	Ch3	Ch4	Ch5	Ch6	Ch7	Ch8
Kneeing start pose	1557	46	74	-1492	4000	-3659	-4000	-1469
	2080	-171	324	207	4000	-1715	-4000	-1534
Kneeing flexion	3753	-375	-267	3	-223	-2103	-611	-561
	3104	-299	-358	85	-272	-3399	-921	-443
Kneeing extension	269	461	200	-482	4000	4000	-4000	-4000
	292	473	186	-446	4000	4000	-4000	-4000
Kneeing normal pose	-1009	809	-63	1185	-790	606	79	299
	-1107	700	-62	1210	-880	645	87	341

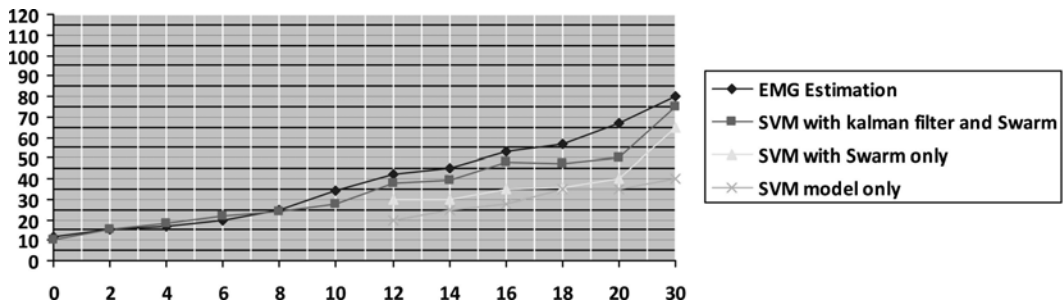


Figure 5. Estimation of knee angle considering mechanical flexibilities in the knee used for each robot kneeling phase error rate 15% due to some more kinematics model for the robot leg must be added to the kalman estimator.

extra feature which is Kalman Filter prediction of Knee-Joint Angel Estimation. As it can illustrated. The last proposed model out performs the first ones. However, still overall accuracy average reaches only 92.3%.

6 CONCLUSION AND FUTURE WORK

Rehabilitation robotics systems can facilitate and help stroke patients to move about almost without restrictions. However, the analysis of real-time EMG sensors in very complex. Also, the selection of accurate neural network classifier is very hard. In this paper, two SVM machines neural networks have been trained and optimized with swarm intelligence with 8 real-time features extracted from 8-channel EMG signal to classify different normal and Auto-aggressive actions. Also, Kalman filter has been used for EMG signal Estimation for better robot control of motions. The hybrid algorithm shows high performance of the application of this algorithm shows better accuracy in most of actions groups. The designed ANN structure has not yet been tested for the EMG signals from disabled people. However, the overall robotics controller, kinematics and dynamics still needs to be further investigated.

REFERENCES

- [1] V. L. Calderita, J. L. Manso, P. Bustos, C. Suárez-Mejías, F. Fernández, and A. Bandera, "THERAPIST: Towards an Autonomous Socially Interactive Robot for Motor and Neurorehabilitation Therapies for Children," *JMIR Rehabilitation and Assistive Technologies (JRAT)*, vol. 1, p. e1, Oct 2014.
- [2] <http://www.europment.org/library/2014/interlaken/bypaper/CSC/CSC-15.pdf>
- [3] Humanoid Team Humboldt: Homepage of Humanoid Team Humboldt, <http://www.humanoidteam-humboldt.de>.
- [4] M. I. Ibrahimy and Md. RezwanaulAhsan, Design and Optimization of Levenberg-Marquardt based Neural Network Classifier for EMG Signals to Identify Hand Motions, *MEASUREMENT SCIENCE REVIEW*, Vol. 13, No. 3, 2013.
- [5] Nahla Farid, Bassant Elbagoury, M. Roushdy and Abdel Badeeh M. Salem, A Comparative Analysis for Support Vector Machines for Stroke Patients, *WSEAS, Recent Advances in Information Science*, ISBN: 978-960-474-304-9, 2013.
- [6] Minas V. Liarokapis and Panagiotis K. Artemiadis, Learning Task Specific Models for Reach to Grasp Movements: Towards EMG-based Teleoperation of Robotic Arm-Hand Systems, 2012.
- [7] S. Mangold, Evidenzbasiertes Arbeiten in der Physio- und Ergotherapie: Reflektiert-systematisch-wissenschaftlich fundiert. Springer-Verlag GmbH Berlin Heidelberg.
- [8] J. Ruhl and V. Laubach, Funktionelles Zirkeltraining: das moderne Sensomotoriktraining für alle. Meyer & Meyer.
- [9] H. Sakakima, K. Ijiri, F. Matsuda, H. Tominaga, T. Biwa, K. Yone, and Y. Sankai, "A newly developed robot suit hybrid assistive limb facilitated walking rehabilitation after spinal surgery for thoracic ossification of the posterior longitudinal ligament: A case report," vol. 2013, pp. 1–4. [Online]. Available: <http://www.hindawi.com/crim/orthopedics/2013/621405/>
- [10] S. B. Godfrey, R. J. Holley, and P. S. Lum, "Clinical effects of using HEXORR (hand exoskeleton rehabilitation robot) for movement therapy in stroke rehabilitation:," vol. 92, no. 11, pp. 947–958. [Online]. Available: <http://content.wkhealth.com/linkback/>
- [11] C. Geroïn, S. Mazzoleni, N. Smania, M. Gandolfi, D. Bonaiuti, G. Gasperini, P. Sale, D. Munari, A. Waldner, R. Spidalieri, F. Bovolenta, A. Picelli, F. Posteraro, F. Molteni, M. Franceschini, and Italian Robotic Neurorehabilitation Research Group.
- [12] Ferdinan Widjaja, Cheng Yap Shee, Win Tun Latt, Wing-Lok Au Kalman filtering of accelerometer and Electromyography (EMG) data in pathological tremor sensing system, in proceedings of IEEE international conference on robotics and automation 2008.
- [13] Alkan, A., Günay, M. Identification of EMG signals using discriminant analysis and SClassifier. *Expert Syst. Appl.* 2012, 39, 44–47.

- [14] A. Subasi, "Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders," vol. 43, no. 5, p. 576586. [Online]. Available: <http://dx.doi.org/10.1016/j.combiomed.2013.01.020>
- [15] S. Wallot, R. Fusaroli, K. Tynl, and E.-M. Jegind, "Using complexity metrics with r-r intervals and BPM heart rate measures," vol. 4. [Online]. Available: [http://www.frontiersin.org/ComputationalPhysiology and Medicine/10.3389/ fphys.2013.00211/abstract](http://www.frontiersin.org/ComputationalPhysiology%20and%20Medicine/10.3389/fphys.2013.00211/abstract)
- [16] C. Hasson, R. Van Emmerik, G. Caldwell, J. Haddad, J. Gagnon, and J. Hamill, "Influence of embedding parameters and noise in center of pressure recurrence quantification analysis," vol. 27, pp. 416–422.
- [17] H. Karl and A. Willig, *Protocols and architectures for wireless sensor networks*. John Wiley & Sons.
- [18] R. Faludi, *Building wireless sensor networks with ZigBee, XBee, Arduino, and processing*. O'Reilly Media.

Solving the problems of linguistically diverse the 1st year university student's using digital learning

D. Ratniece

Riga Technical University, Riga, Latvia
University of Liepaja, Liepaja, Latvia

S. Cakula

Vidzeme Univeristy of Applied Sciences, Valmiera, Latvia

ABSTRACT: University students from diverse language backgrounds encounter some difficulty every day. Often tuition language and culture are different from what they have learned at home. In this article, the term linguistically diverse students will be used to refer to “students whose first language is other language than Latvian. The scientific data collected by the author are from 2013/2014 academic year until 2015/2016 academic year (3 academic years). Machine translation technology is constantly being applied by linguistically diverse students, but a machine cannot assess whether a sentence sounds good or bad. Motivation is a powerful force in second-language learning. The aim of the research is to find out how to solve linguistically diverse 1st year university students' problems using digital technologies and digital learning. Educational support and motivation enhancement are very important. Various courses that use digital learning can be applied as the spectrometers which help to identify and solve the 1st year students' education quality problems and increase learning motivation.

Keywords: linguistically diverse students, digital learning, machine translation, motivation

1 INTRODUCTION

1.1 *Language*

Language can be defined as a means of communication that shapes cultural and personal identity and socializes one into a cultural group (Goillnick & Chinn, 2006). It is impossible to separate language and culture. University students from diverse language backgrounds encounter some difficulty every day. Because language and culture are so intertwined, language minority students are expected to learn and use a tuition language and new cultural dispositions effectively. Often tuition language and culture are different from what they have learned at home. In this article, the term linguistically diverse students will be used to refer to students whose first language (L1) is other than Latvian. State universities tuition language is Latvian.

Data from 2013/2014 academic year until 2015/2016 academic year (3 academic years) indicate that the 1-st year Riga Technical University Engineering Telecommunications faculty linguistically diverse students comprise approximately 40% of all students (Ratniece, Cakula, 2015). Therefore, academic personnel must be aware of diversity in their classrooms and how it may have an impact on students' achievements. (The

Bachelor study programme provides a blend of knowledge from electrical engineering and computer science focusing on communications networks and systems, encoding theory, information/optical processing and transmission. The aim of the programme is to provide an academic education and prepare students for further studies at the Masters level).

1.2 *Digital learning*

Digital learning is any instructional practice that effectively uses technology to strengthen a student's learning experience. It emphasizes high-quality instruction and provides access to challenging content, feedback through formative assessment, opportunities for learning anytime and anywhere, and individualized instruction to ensure all students reach their full potential to succeed in education and a career. Digital learning encompasses many different facets, tools, and applications to support and empower teachers and students, including online courses, blended or hybrid learning, or digital content and resources. Additionally, digital learning can be used for professional learning opportunities for teachers and to provide personalized learning experiences for students.

1.3 *Machine translation*

Machine translation technology as digital learning tool is constantly being applied by linguistically diverse students in the study process, but a machine cannot assess whether a sentence sounds good or bad. A machine is also incapable of managing nuances, subtexts, symbolism or wordplay; it cannot control mood or tone. So, machines are not likely to replace human translators in near future.

Here's what a machine translator can do:

- Like a bilingual dictionary, it can match a word in one language with a word in another language. However, the same word may have different meanings. For example, “spirits” can be either souls or alcoholic drinks in English.
- When it has to choose between different possible translations, a machine translator can make statistical “guesses” at the context. For example, in a sentence that talks about both meat and spirit, the machine might guess that the word “spirit” refers to alcohol.

However, the problem with machine translation is that a machine is only a machine. It matches components and follows rules. It doesn't actually know what it's talking about. A machine cannot assess whether a sentence sounds good or bad. A machine is also incapable of managing nuances, subtexts, symbolism or wordplay; it cannot control mood or tone.

1.4 *Motivation*

Motivation is a powerful force in second-language learning. Motivation governs a need to communicate, to make friends, to identify with a social group, to become part of a community and to begin to plan one's future.

1.5 *Aim, questions, research methods*

Aim of the research:

To identify the problems and motivation factors for linguistically diverse 1st-year university students caused by quality of digital learning.

Questions of the research:

- The use of digital learning in itself does not constitute an enhancement of the quality of teaching and learning, but it is a potential enabler for such enhancement;
- Evaluation of linguistically diverse 1st-year universities students' work using digital learning (many different facets, tools, applications and online course management system) motivates them to study.

Research methods. As the theoretical framework, the following methods were applied:

- Evaluation of all homework assignments uploaded in the e-learning environment.
- The risk-taking assessment by Schubert's method;
- Diagnostics of the person on motivation by T. Elersa methods;
- Failure avoidance motivation method;
- A survey with the assessment of the course “Entrepreneurship (Distance Learning e-course)”, developed by the author of this article.

2 THEORETICAL BACKGROUND

Dulay, Bust, and Krashen (1982) in their survey of major findings in second-language research indicate that the most beneficial environment for the learner is one which encourages language learning in natural surroundings for genuine communication. Furthermore, it has been shown that optimal second-language (L2) learning takes place in an environment:

- which is non-threatening, in which the learner feels free to take chances and make mistakes.
- which is linguistically and non-linguistically diverse (i.e., no grammatically sequenced syllables, no attempt to homogenize the environment so that learners understand everything).
- in which learners focus on tasks and activities of interest to them, and use language as a tool to get things done (i.e., very little explicit discussion of language).
- in which learners' interests and needs serve as the basis for learning activities.
- in which learners' talk is considered to be the task—as in “being on task”: small talk, jive, and tall tales are not only tolerated, they are encouraged, and not just at “sharing time” but throughout the day.

To become life-long language users, L2 students as well as native-Latvian speaking students need to gain control over language and feel comfortable about using the language. The ensuing principles for second-language instruction can help lecturer create supportive language environments:

Latvian as a Second Language (LSL) students' learnings should be built on the educational and personal experiences they bring to educational establishment. In language learning, students should be encouraged to use their previous experiences with oral and written language to develop their second language and to promote their growth to literacy. Students bring to educational establishment cultural identities, knowledge, and experiences that should be awarded by instructional practices rather than replaced or forgotten as learning takes place (Au and Jordon, 1981; Hudelson, 1986; Edelsky, 1986; Cummins, and Swain, 1986;

Enright and McCloskey, 1988; Cummins, 1986; Jordan, 1985; Diaz and Moll, 1987). Socializing, learning, questioning, and wondering are some of the many things that one is able to do when one learns a language. However, these things are not quickly learned; it takes many years to develop full-fledged competence (Wong-Fillmore, 1983; Collier, 1987). Furthermore, rates of development of oral proficiency vary considerably in LSL students. Consequently, lecturers, not just LSL specialists, need to address the learning needs of LSL students and adjust their instruction accordingly to meet the different levels of Latvian proficiency, different learning rates, and styles of their students. Instructional convenience does not mean, however, a “watered-down” curriculum.

It has long been recognized that if LSL students are to ‘catch-up’ or ‘keep-up’ with their native-Latvian speaking peers, their cognitive and academic growth should continue while the second language is developing. Thematic units (as opposed to exercises in grammatical structures), where language is integrated with academic content, appears to be an effective way to simultaneously develop students’ language, subject area knowledge, and thinking skills. Thematic units help involve students in real language use—use of language interactively across a variety of situations, modes, and text types. Digital learning and machine translation technology is constantly being applied by linguistically diverse students in the study process and that strengthens a student’s learning experience and motivation.

3 RESEARCH METHODOLOGY

Methods for diagnostics of the degree of risk preparedness, motivation to success and motivation to avoiding failures have been tested for all students many times during course: at the beginning, in the middle of course and at the end.

3.1 *Data collection methodology*

Motivation becomes a positive force. Anxiety becomes an inhibitor. Self-confidence is very much related to second-language learning as is a low anxiety level and a tendency to be risk-takers and do guess work. As the student becomes more secure in the second language, it is entirely likely that the native language precedes, to some extent. As vocabulary in the second language increases, words in L1 may well be forgotten. During the second-language learning process, a learner may insert words from each language in the same sentence. Again, this tendency demonstrates a motivation to speak the second language and is a way of

permitting precise expressions which carry cultural content and can be stated in a given language.

Web-based e-learning platforms allow educators to construct effective on-line learning study courses by uploading various categories of study materials. E-learning platform allows usage of a wide range of on-line learning tools as forums, discussion forums, e-mail messaging, as well as combining face-to-face and on-line approaches. The purpose of these technologies is: to deliver study materials to a student, improve students’ skills, assess skills and knowledge, and achieve better learning outcomes. Fast and immediate feedback is possible. In e-learning platforms produce data logging. Logged data can be used for later analysis. There are two types of data:

- Data produced by students or teachers and represent the content of the learning course;
- Data made by the system based on student’s activities like in system-spent time, kept sessions, the number of clicks on items of the content, etc. (Ratniece, Cakula, Kapenieks, Zagorskis).

At Riga Technical University (RTU), e-learning platform MOODLE has been maintained. In the period from October until December of academic year 2013/2014, academic year 2014/2015 and the academic year 2015/2016 the course “Entrepreneurship (Distance Learning e-course)” to 1st-year students was provided. The course was conducted by RTU Professor A. Kapenieks. The author, as the Assistant to the Professor, supplemented the lecture content. Author’s study, entitled “Use of Social microblogging to motivate young people (NEETs) to participate in distance education”, was presented.

3.2 *Description of the experiment*

The research was carried out during the lectures and the final exam of a course “Entrepreneurship (Distance Learning e-course)” with the 1st year students (respondents) participating on a voluntary basis. Home works—the course had two homework assignments:

- «The search of Business Ideas on the Internet»;
- «Your business idea».

The author has evaluated all homework assignments uploaded in the e-learning environment. Reviews and comments were added, with the aim to encourage and motivate students to prepare their business plans in time and of good quality. Each comment was prepared according to the results of the content analysis. The feedback comments to students were written in a positive, supportive and motivational manner, personally addressing each student. The author has noted on students’ written

language mistakes. The assessment of the student's homework was done by the author concerning seven criteria:

1. Actuality or viability of idea;
2. Technological solution or how to enforce;
3. Marketing—promotion of goods or services in the market;
4. Competition;
5. Financial security (e.g., planned revenues, expenses, financial support for the company's start-up and ongoing development (bank loan, other resources etc.);
6. The ability of a company to realize the idea;
7. The potential risks.

4 DATA ANALYSES AND EVALUATION

Every year the Latvian and linguistically diverse students' success rate increases. Looking at the two groups of proportion, it is evident that linguistically diverse students total proportion increases, which have successfully completed a course "Entrepreneurship (Distance Learning e-course)" (Table 1).

Home works' (uploaded in the e-learning environment) quality level of linguistically diverse 1st-year university students every next year became better concerning the use of machine translator.

Data relating to the second homework "Your business idea" of a course "Entrepreneurship (Distance Learning e-course)" shows the real-study example of machine translation, linguistically diverse 1st-year universities students took and used online translator on Google, to translate it to Latvian. The topic concerning linguistically diverse

1st year university students' translations using machine translation (online translator on Google) is included in the table 3 (see the points—5,6,7,8), as well as in the table 4 (see the point 2.3. and the point 2.4.).

Table 2. Students' evaluation of the effectiveness of the form practiced in the course "Entrepreneurship (Distance learning course)" (academic years 2013/2014th and 2014/2015th).

All respondents (Latvian and linguistically diverse 1st year university students) indicate that e-learning and traditional forms of study need to be kept in balance, because e-learning provides a great advantage to learn anywhere, anytime. A successful guidance through the study process, however, is just as important, and can only be ensured when a teacher is present.

It should be noted that in the academic year 2013/2014 teacher comments on-line environment (including the correct use of the Latvian language) were written in assessing students' second home work.

In the academic year 2014/2015 teacher comments on-line environment were written in assessing students' first and second home work.

In the academic year 2015/2016 teacher comments on-line environment were written in assessing students' first home work.

Students' average assessment shows that students appreciated the teacher's job better when the teacher had evaluated both homework assignments.

The number of students has been decreased over the period of one semester: only two-thirds of all students turned in the second assignment. The author has identified that there exist some certain

Table 1. Latvian and linguistically diverse students' results over the 3 academic years.

Academic year	Students' two groups in relation to the language use	The number of students who started a course "Entrepreneurship (Distance Learning e-course)"		The second homework "Your business idea"		The teacher's comments according to the incorrect Latvian language use		The number of students, who successfully completed course	
		Every group	Total	Every group	Total	Every group	Total	Every group	Total
		2013/2014	Latvians	80	142	62	84	3	21
	Linguistically diverse	62		22		18		30	
2014/2015	Latvians	78	142	43	80	2	13	53	79
	Linguistically diverse	64		37		11		26	
2015/2016	Latvians	77	132	32	64	0	4	32	56
	Linguistically diverse	55		33		4		24	
TOTAL		416		228		38		242	

Table 2. Students' evaluation of the effectiveness of the form practiced in the course "Entrepreneurship (Distance learning course)" (academic years 2013/2014th and 2014/2015th).

Form of study	Low rating			Average rating				High rating			Weighted Mean
	1	2	3	4	5	6	7	8	9	10	
1. Lectures 2013/2014				3	6	16	34	29	11	6	7,30
2. Lectures 2014/2015	1	1	2	1	1	10	25	18	12	8	7,80
2. Discussions 2013/2014			2	3	5	13	18	35	21	8	7,73
3. Discussions 2014/2015		1	1	1	1	3	13	24	18	17	8,40
4. Insertion in ORTUS 2013/2014			1		7	11	19	35	15	17	7,90
5. Insertion in ORTUS 2014/2015			1	3	4	5	15	20	15	16	8,01
6. Teachers' comments 2013/2014			1	2	7	12	15	33	17	18	7,88
7. Teachers' comments 2014/2015	1		1	2	4	4	9	14	12	31	8,54

Table 3. Students' evaluation of the effectiveness of the form practiced in the course "Entrepreneurship (Distance learning course)" (2015/2016).

1. Traditional forms of study	Low rating			Average rating				High rating			Weighted Mean
	1	2	3	4	5	6	7	8	9	10	
1. Lectures				1		3	18	16	10	8	7,96
2. Discussions					1	2	7	16	19	11	8,09
2. Digital opportunities to increase students' education quality and motivation											
2.1. Digital environment interface (cumbersome—easy to learn)				1	1	5	8	15	15	11	8,21
2.2. Digital accessibility (not at all—in full)					1	3	5	8	18	21	8,82
2.3. Assignment preparation and insertion in ORTUS system (unsuccessful—successful)			1			1	5	13	11	25	8,88
2.4. Teachers' comments in ORTUS system (redundant—needed)	1	1		3	3		4	15	11	18	8,14
2.5. Functional content of the course (inadequate—adequate)				1	1	2	6	19	11	16	8,64
2.6. Content structuring (opaque—transparent)			1	1	1	4	5	13	18	13	8,32
2.7. Motivation tests—improve the study process (unimportant—important)			1	1	1	7	6	8	11	10	7,57

"risk factor" level for student to be dropped out of the course that correlate with a student's level of preparedness, readiness, and eagerness.

In 2013/2014 academic year 107 questionnaires were issued and filled in by students, 107 students had finished the course. In 2014/2015 academic year 79 questionnaires were filled in, 77 students had finished the course. In 2015/2016 academic year 56 questionnaires were filled in, 53 students had finished the courses.

Testing mean of activity between native language speaking student group and linguistically diverse students there are difference on 95% probability. Linguistically diverse students are more active

(mean of activity for native speaking students is 159 and for linguistically diverse students is 246). There are no difference on course evaluation and final grade between both groups on probability level 95%.

Also all means of psychological test results are equal for both student groups on probability level 95% (see Table 4).

For example: H₀: the mean of final course evaluation are equal between native language student group and linguistically diverse student group.

H_a: the mean of final course evaluation are not equal between native language student group and linguistically diverse student group.

Table 4. Independent samples test.

	T	df	Sig. (2-tailed)	Mean Differ.
Activity	2,899	115	,004	-86,846
1_st_practical_evaluation	2,042	115	,043	-1,519
2_nd_practical_evaluation	1,257	115	,211	-1,033
Final	-0,779	264	,437	-1,777
Motivation to avoiding failures at the beginning	0,845	139	,399	0,577
Motivation on success in the beginning	1,644	159	,102	-0,564
Diagnostics of the degree of risk preparedness in the beginning	-0,183	153	,855	-0,366
Motivation to avoiding failures at the end	0,922	75	,359	0,902
Motivation on success at the end	0,137	84	,891	0,079
Diagnostics of the degree of risk preparedness at the end	-0,314	75	,754	-1,010

T-test value -0,779 not increase critical value on significal level 0,05—it means that H0 is right. The average evaluation between both student groups are equal on probability level 95%.

5 CONCLUSIONS AND FUTURE WORK

Digital learning and machine translation technology emphasizes high-quality instruction and provides access to challenging content, feedback through formative assessment, opportunities for learning anytime and anywhere, and individualized instruction to ensure all students reach their full potential to succeed in education and a career. It confirms that:

1. Linguistically diverse student's homework evaluation of the on-line environment improves students' knowledge, as well as learning the language, if the teacher points to the grammatical errors resulting from the use of machine translation.
2. There will be an increase at higher education establishments concerning the linguistically diverse student's total proportion.
3. Homework assignments' (uploaded in the e-learning environment) quality level of linguistically diverse 1st-year university students concerning the use of machine translator became better year after year.

Digital learning objects are often considered complete and whole the moment they are uploaded into a digital learning repository. However, these may be versioned over time, and they should be labeled as such. Updates to digital learning objects should generally be done in the following contexts: when the paradigm has shifted in a field; when critical data has changed; when there are important policy changes; when new learning experiences

and methods are possible to enhance the learning. The learning objects that are hosted on sites may be continually updated and revised for quality. It is important to notify an installed base of users of updates to materials if they choose to receive such notifications.

REFERENCES

1. Gollnick, D.M., & Chinn, P.C. (2006). *Multicultural education in a pluralist society* (7th ed.). Upper Saddle River, NJ: Pearson.
2. Ratniece, D., & Cakula, S. (2015). Digital Opportunities for Student's Motivational Enhancement. In: *Procedia Computer Science*, 2015, Vol.: International Conference on Communication, Management and Information Technology, pp. 22–22, Prague, 2015.
3. Ratniece, D., Cakula, S., Kapenieks, K., & Zagorskis, V. Digital Opportunities for 1-st Year University Students' Educational Support and Motivational Enhancement. In: *The 1-st International Conference on Advanced Intelligent Systems and Informatics (AISII2015) November, 28–30, 2015, Beni Suef, Egypt, Recent Research on Advanced Intelligent Systems and Informatics, Egypt, Beni Suef, November, 28–30, 2015. Beni Suef: Springer International Publishing, 2015, pp. 1–10, e-ISBN 978-3-319-26690-9. ISSN 2194-5357, Available from: doi: 10.1007/978-3-319-26690-9 (included in SCOPUS).*
4. Dulay, H., M. Burt, & S. Krashen. (1982). *Language Two*. New York: Oxford Press.
5. Au, K., & C. Jordan. (1981). "Teaching Reading to Hawaiian Children: Finding a Culturally Appropriate Solution." In *Culture and the Bilingual Classroom: Studies in Classroom and Ethnography*, edited by K. Au, G. Guthrie, and H. Trueba. (pp. 139–152). Rowley, Mass.: Newbury House Publishers.
6. Hudelson, S. (1986). "ESL Children's Writing: What We've Learned, What We're Learning." In *Children and ESL: Integrating Perspectives*, edited by V. Allen and P. Rigg. (pp. 23–54). Washington, D.C.: Teachers of English to Speakers of Other Languages.

7. Edelsky, C. (1986). *Writing in Bilingual Program: Habia Una Vez*. Norwood, N.J.: Ablex Publishing Corporation.
8. Cummins, J., & M. Swain. (1986). *Bilingualism in Education: Aspects of Theory, Research and Policy*. London: Longman.
9. Enright, D., & M. McCloskey. (1988). *Integrating English: Developing English Language and Literacy in the Multilingual Classroom*. Reading, Mass.: Addison-Wesley.
10. Cummins, J. (1986). "Empowering Minority Students: A Framework for Intervention." *Harvard Educational Review* 56: 18–36.
11. Jordan, C. (1985). "Translating Culture: From Ethnographic Information to Educational Program." *Anthropology and Education Quarterly* 16: 105–123.
12. Diaz, R., & L. Moll. (1987). "Teaching Writing as Communication: The Use of Ethnographic Findings in Classroom Practice." In *Literacy and Schooling*, edited by D. Bloome. (pp. 195–221). Norwood, N.J.: Ablex Publishing Corporation.
13. Wong-Fillmore, L. (1983). "The Language Learner as an Individual: Implications of Research on Individual Difference in the ESL Teacher." In *On TESOL '82: Pacific Perspectives on Language Learning and Teaching*, edited by M. Clarke and J. Handscombe. (pp. 157–173). Washington, D.C.: Teachers of English to Speakers of Other Languages.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Multi-agents based framework for selecting cloud service provider

Mohamed Abo-Rizka

Centre of Excellence, Arab Academy for Science and Technology, Cairo, Egypt

Radwa El-Awadi

Computer and Information System, Sadat Academy for Management Sciences, Cairo, Egypt

ABSTRACT: Cloud Services become increasingly popular among enterprises, developers, and organizations. Cloud Service Customers (CSCs) can easily spin up hundreds of Virtual Machines (VMs) and limit their payment for only what they can actually use without any up-front investment. This advantage has motivated more and more organizations including governments, schools, enterprises, and content providers, to migrate their applications to the Cloud. Performance in Cloud Computing is defined through Service Level Agreement (SLA) contracts usually between two main agents. These agents represent the CSPs and CSCs. Under the Cloud based agents' approach; these agents usually with different requirements, exchange messages to negotiate Service Level Agreement for a Cloud Computing service, however both could fail to understand the correct meaning of their messages and the offered service quality, due to the differences existing within their inner needs. In this case, an automatic negotiation is required to mutually make the messages exchanged about the Cloud service characteristics understandable. Our previous work introduced an approach assisting cloud customers in selecting the most reliable service offered by the CSP. In this work, a proposed framework explaining the automatic negotiation process between Cloud agents as well as integrating the roles of these agents into a one Trusted Third Party (TTP) for Infrastructure Monitoring Resource Providers' (IMRP) and Recommending System (RS) will be presented.

Keywords: Cloud Services, Cloud agents, CSP, CSC, SLA, automatic negotiation, TTP, IMRP

1 INTRODUCTION

“Cloud Computing is a new trend in IT field where the computing resources are delivered as a service. These computing resources are offered as pay-as-you-go plans and hence have become attractive to be cost effective for customers rather than the traditional infrastructures”. It was defined in El-Awadi and Abu-Rizka (2015) and Habib et al., (2012). With the growing number of CSPs offering a variety of on-demand services, the process of adopting cloud services and selecting a reliable CSP is becoming time consuming involving complex and lengthy negotiations to mediate user requirements with those of the Cloud SLAs. To make the method more efficient, there is a need for an effective, dynamic, and flexible automated approach on negotiation for resolving conflicts and mediating user requirements in SLAs.

According to Yaqub et al., (2011) SLA refers to the contractual obligations between a service customer and a service provider, representing guarantees of Quality of Service (QoS) requirements which are defined in SLA.

Automatic SLA Negotiation process has to contain; a) set of services which the CSP will deliver with a complete and specific definition for each service, b) the automatic terms of the agreement should be clearly defined and able to anticipate future problems e.g. (Upgrading Services) c) set of Quality of Service (QoS) metrics in Messina et al., (2014) to measure whether the provider is offering the services as guaranteed, d) an auditing mechanism to monitor the QoS and e) the remedies available to customer and provider if the terms are not satisfied.

The automatic establishment of SLA requires precise and unambiguous definition of the agreement as well as customizable engines to support automated negotiation on the details of the agreement for both contracted parties.

In the common practice, CSPs publish the characteristics of the offered Cloud services in a way that makes a comparison a boring and tedious task. Consequently, the customer has to perform complex and boring converse tasks to compare different deals for cloud services.

Additionally, from the obstacles facing CSCs which were revealed from the previous work, some

CSCs had overestimated the required resources, an issue that had led to the existence of underutilized resources.

In the current work, it was found necessary to propose a new framework for monitoring the utilization behavior of CSCs with similar needs through a one TTP for IMRP and RS.

The paper is organized as follows: Related work is described in Section 2. In Section 3 the Cloud SLA KPIs are clarified as well as the novelty of automated SLA negotiation mechanism through Cloud agents to provision an automated establishment of SLAs that enhances the utility of SLAs for both CSPs and CSCs are illustrated in section 4. Section 5 deals in detail with the proposed framework and its components, while Section 6 will perceive the experimental evaluation. Section 7 draws some final conclusions and future work.

2 RELATED WORK

Due to the increased range of the CSPs, the necessity of SLA, SLA management, negotiation and Service Level Objectives (SLOs) monitoring has also heavily increased.

Recent research has focused on Cloud SLA negotiation process, contract management, scalable monitoring system for clouds and cloud assessment and recommendation system.

In El-Awadi and Abu-Rizka (2015), a framework for negotiating Service Level Agreement of Cloud based Services was proposed. Such a framework intended to empower the customers to select among the different CSPs service offerings. The evaluation of this framework showed that underutilized resources existed as the customers lacked the experience in selecting the adequate resources matching their needs. This had actually led to missing the main advantage of Cloud Computing which is pay per use.

Actually, Brinkmann et al., (2013), had previously pointed out the necessity of monitoring the infrastructure data supported by suppliers which are organized in a distributed and easily scalable tree structure.

Additionally Longo et al., (2015), proposed the extension of Web Service Level Agreement (WSLA) for modeling the contracts and supporting contract parties during the service composition and monitoring.

As well as Frey et al., (2013) proposed selected KPIs for Cloud SLAs that have to be monitored and described possible Service Level Objectives (SLOs).

Besides Son and Jun (2013) designed a Cloud SLA negotiation mechanism for interactive and flexible SLA establishment.

Also Alemeye and Getahun (2015), proposed a Cloud readiness assessment framework and an expert system that assessed Cloud readiness and recommend which service model to adopt.

Aceto et al., (2013) provided a survey on Cloud monitoring which analyzed motivations for Cloud monitoring and discussed the properties of a monitoring system for the Cloud.

Likewise Brinkmann et al., (2013) presented the scalability approach for Cloud monitoring system that must scale well without wasting resources. We adopted this approach so as to monitor the infrastructure resources that are provided from CSPs in order to assist the CSCs to fully utilize the required resources and get benefit from the Cloud advantage of paying per request.

Based on the previous research, this framework intends to come up with an integrated framework merging the roles of cloud agents for the negotiation process and SLA establishment into a one Trusted Third Party for Infrastructure Monitoring Resources Providers and Recommending System.

Based on the previous research, this framework intends to come up with an integrated framework merging the roles of Cloud agents for the negotiation process and SLA establishment into a one Trusted Third Party for Infrastructure Monitoring Resources Providers and Service Selection.

3 CLOUD SLA KPIs

Service Level Agreements (SLAs) indicate the guaranteed and the expected performance characteristics between service CSPs and CSCs. Consequently, all significant and relevant information and services are established. The most vital part of a SLA is the accurate depiction of the QoS requirements and the service KPIs according to (Longo et al., 2015).

The following section illustrates the QoS requirements, the service KPIs and structure of Service Level Agreements.

The establishment of SLAs provides certain requirements to CSCs and CSPs. CSCs need to be able to meet their certain requirements in order to successfully define SLAs, which are discussed in Wu et al., (2015) and listed below.

CSCs have to a) perceive the obligations and responsibilities that are controlled by SLA. b) Be able to clearly define precisely the services that are regulated by SLA. c) Define the ways by which SLA KPIs can be measured and monitored. d) Know the level of service performance based on the characteristics of the services. (Longo et al., 2015).

These requirements are vital so that the CSCs can add in the right SLAs KPIs values, and to

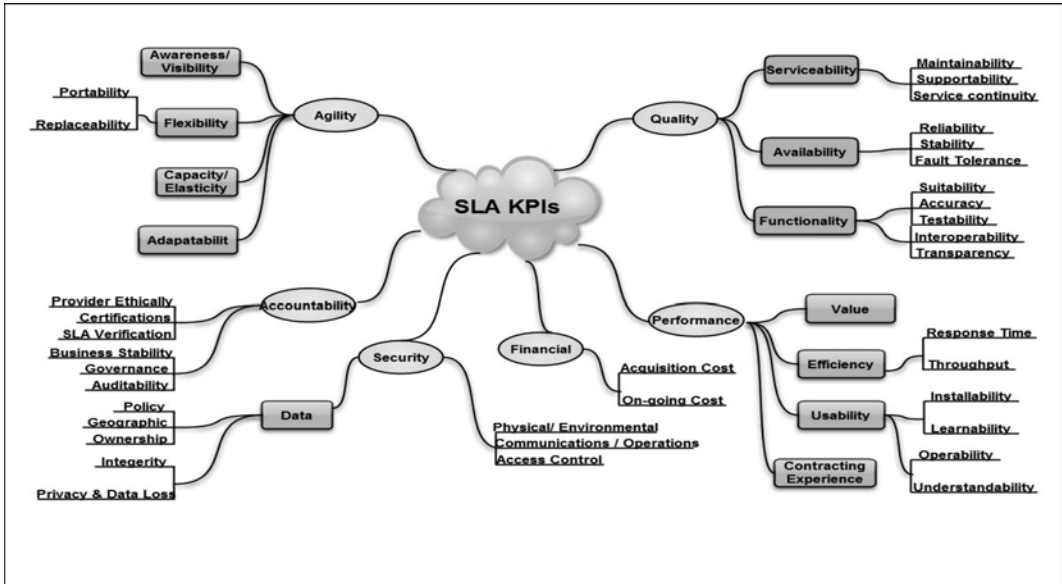


Figure 1. Cloud service KPIs.

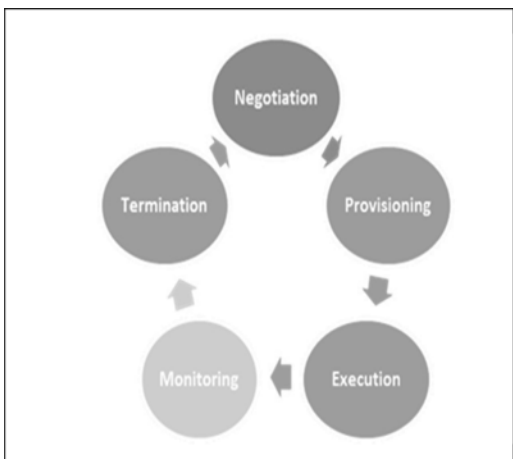


Figure 2. Cloud service KPIs.

comprehend implications of their choices. (Frey et al., 2013) Moreover, a SLA ought to satisfy the complementary tasks: 1) enforce penalties when service requirements are not met. 2) depict clearly a service so that the CSCs can easily comprehend the operation of the services. 3) Indicate the service quality to be given in refined element. 4) Describe completely the key performance indicators KPIs, metrics and service levels as shown in Fig. 1. 5) Breakdown straightforwardly all of the expenses.

SLA has a life cycle that comprises several phases for a successful use of SLA. (Frey et al.,

2013) There are diverse perspectives on whether the definition phase of the SLA is one of its life cycle or not, since this can likewise be considered as a part of the preconditions.

Figure 2 shows the SLA life cycle. The individual phases are briefly showed:

The prerequisites of SLA life cycle is the definition of primary SLA template based on which the negotiation phase is started. In the negotiation phase, the deliverable service KPIs is negotiated with the CSP. However; in the provisioning phase, the input to potency of the agreement is striking by the signatures of both agents. At this time, the provided services are provisioned. During the execution phase the customer utilizes the service according to his views. Equivalent to this, the monitoring phase the runtime service is checked and assessed contrary to the service SLA. If desirable, remedial actions are and reports and are made for the accomplices. The final termination phase donates the end of the CSCs utilization and starts retire of the service.

4 AUTOMATED SLA NEGOTIATION

In El-Awadi and Abu-Rizka (2015), introduced an approach that exhibited main four Cloud agents scenarios.

The first scenario as shown in Fig. 3, CSPs are accountable for making a service available to interested parties. It attains and manages the computing

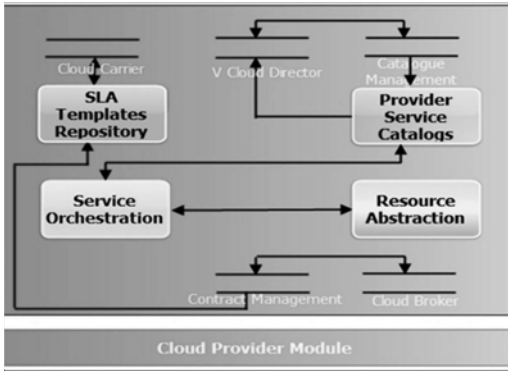


Figure 3. CSP agent module.

infrastructure required for providing the services, runs the Cloud software that provides the services, and makes procedure to convey the Cloud services to the CSCs through network access. CSP conducts its activities through a service orchestration.

The CSP agent comprises the following components; Resource Abstraction, Service Orchestration, SLA Templates Repository and Provider Service Catalogues. The CSP agent normally communicates with other parties like Cloud Carrier, V Cloud Director, Catalogue Management, Contract Management, and Cloud Broker. These components were discussed in details in El-Awadi and Abu-Rizka (2015).

The CSC comprises the following components and communicates with other parties as follows in the Fig. 4

As discussed earlier, the complete SLA management over the lifecycle of SLA incorporates observing procedure with a specific end goal to make the SLA arrangement in the middle of CSS and CSP. Figure 5 will be illustrated the components of trusted Cloud Service Auditor (CSA) interfaces with them.

After listing the roles of different Cloud agents, the selection process will be deployed as shown in Fig. 6, the selection process can be fulfilled by observing how the component parties interface with it and the SLA establishment between the CSP and CSC.

The Cloud Service Broker (CSB) agent is representing the trusted third party who establishes a SLA between CSP and SLA Carrier in order to transport Cloud services to the CSCs.

The negotiation process includes the Negotiation Policy Specification and the Negotiation Protocols. The negotiation policy specifications used to identify QoS KPIs. The negotiation protocol refers to a set of rules, steps or sequences during the negotiation process, aiming at SLA establishment.

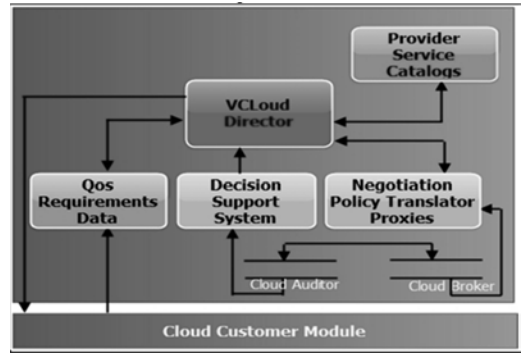


Figure 4. CSC agent module.

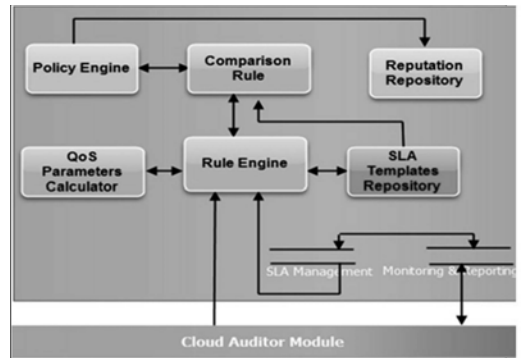


Figure 5. CSA agent module.

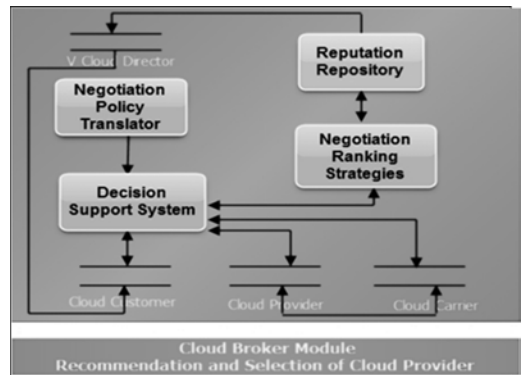


Figure 6. CSB agent module.

If the CSB makes CSCs fill in the negotiation KPIs in the V Cloud Director, CSCs might worry that their QoS requirements would be uncovered to the adversary, who is a broker for the benefit of CSPs services. Accordingly, a creator of

broker needs to decide the deployment position of negotiation process.

5 PROPOSED FRAMEWORK OF TTP FOR IMRP AND SERVICE SELECTION

The proposed framework incorporates negotiation and secure monitoring mechanism involving a TTP for IMRP and RS is developed.

Cloud monitoring can give information about parts of system performance, execution, conduct, advancement, and so forth. The way this information is comprehended, broke down and utilized depends not just on what level of the system is being observed.

In an ordinary IaaS Cloud, CSCs can screen the state of their virtual machine cases so as to think about system load, memory utilization and execution performance. In the same IaaS context, the CSPs would need to screen all VM cases, consistently ensuring SLA confinements are fulfilled. The CSPs would likewise require observing data from the server level, so as to viably control general system load, VM assignment and movement, and so forth. In this way, the purpose of perspective of the element that acquires the observing information (CSCs, TTP, CSP, and so on.) and its part in the framework figure out what sort of data must be given. Distinctive substances require diverse observing information and have diverse visions of the Cloud.

From a general point of view Montes et al., (2013) two fundamental Cloud monitoring visions can be recognized:

CSCs side monitoring vision: From this perspective, the Cloud is viewed as a unique element, equipped for giving a particular arrangement of computational service. Checking information of this sort gives a conceptual depiction of the Cloud service, communicated in the same terms as the service provisioning relationship is built up between the CSC and the CSP through SLA. This observing information offers the CSC to comprehend the attributes of the service, some assistance with receiving and enhances their utilization.

CSPs monitoring vision: From this perspective, the Cloud is viewed as a complex conveyed base, with numerous equipment programming components joined together to give a particular arrangement of service

Monitoring information of this sort gives the CSP learning about the working of the distinctive cloud components, its state, execution, and so forth. This information serves as an inward status control keeping in mind the end goal to ensure SLAs and other administration confinements. It can be likewise utilized as conduct and execution log, to advance system management and utilization of resources.

Figure 7 shows the main components of the proposed framework which are illustrated as follows:

The IaaS CSPs layer is divided into the physical resources and the virtual resources.

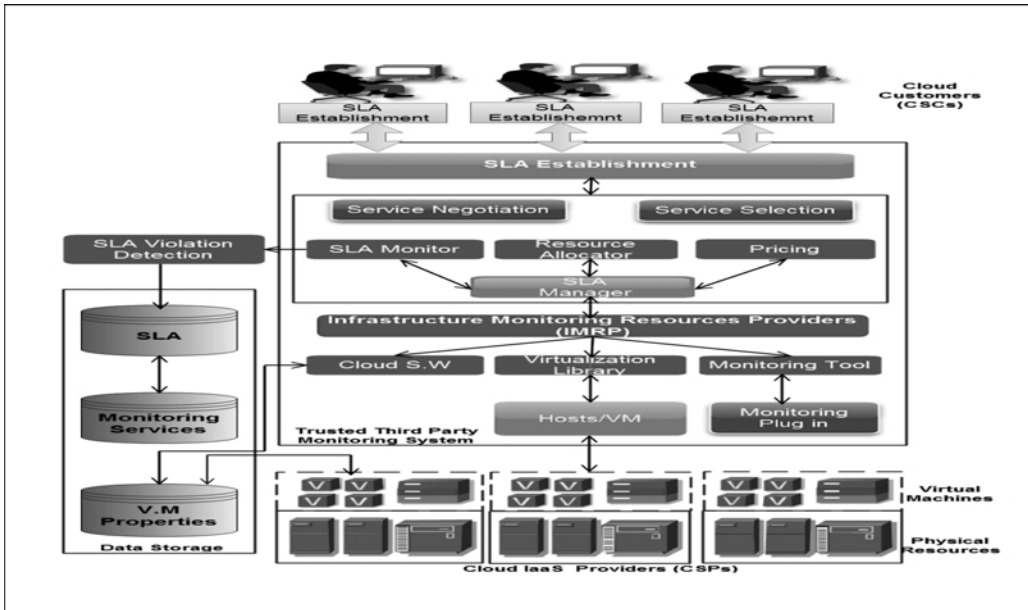


Figure 7. Proposed framework of TTP for IMRP and service selection.

The physical resources are consisting of:

- *Compute*: This is a gathering of all CPU capacities. Basically all data-center servers, either to support or really running a workload, are all part of this compute portion. Compute pool states to the aggregate limit for executing code and running instances. The process to build compute pool is to first inventory all servers and recognize virtualization candidates followed by server virtualization. Furthermore, it is never too shortly to acquaint a framework for management this resource in my perspective is a key speculation and a basic segment for all Cloud services.
- *Networks*: The physical and logical objects putting in place to connect resources and gathered in the network pool. Networking enables resources becoming observable and hence possibly manageable. Networking in cloud computing is more than just remote access, but empowerment for a user to self-serve and consume resources anytime anywhere with any device.
- *Storage*: An enterprise storage solution frequently characterizes as a high cost item with a significant financial and contractual commitment, specialized hardware, proprietary API and software, a dependency on direct vendor support, etc. In cloud computing, storage has become even more noticeable since the ability to grow and shrink based on demands, i.e. elasticity; demand an enterprise-level, reliable, and robust storage solution at a global scale. While enterprise IT is consolidating resources and transforming existing establishment into a cloud computing environment, how to leverage existing storage devices from various vendors and integrate them with the next generation storage solutions are among the high priorities for modernizing a data-center.
- *Memory*: Memory in Cloud Computing represents the capacity of data and information in the Random Access Memory (RAM) of dedicated servers as opposed to relational databases working on similarly slow disk drives. Cloud memory offers business customers, some assistance with including retailers, banks and utilities, to rapidly recognize and detect patterns, analyze massive data volumes on the Cloud, and perform their operations quickly. The drop in memory costs in the present business sector is a main consideration that contributes with the increasing in Cloud memory. This has made in-memory Computing temperate among a wide assortment of utilizations.

The virtual layer represents mainly server virtualization which abstracts physical resources, such as compute, storage, memory and network, to function as logical resources.

Virtual IT resources consist of: VMs, virtual volumes, and virtual networks. VM network components such as virtual switches and virtual NICs

Virtual IT resources gain capacities such as CPU cycles, memory, network bandwidth, and storage space from the resource pools. Virtual networks are defined using network identifiers such as VLAN IDs and VSAN IDs from the respective identity pools. MAC addresses are assigned to virtual NICs from the MAC address pool. It creates an abstraction layer to hide the physical characteristics of resources from users.

In compute system virtualization, a physical machine appears as multiple logical machines (virtual machines), each running an operating system concurrently.

By consolidating IT resources using virtualization techniques, organizations can optimize their infrastructure utilization. By improving the utilization of IT resources, organizations can reduce the costs associated with purchasing new hardware. They also reduce space and energy costs associated with maintaining the resources. Moreover, less people are required to manage these resources, which further lower the cost.

Virtual resources are created using software that enables faster deployment, compared to deploying physical resources. Virtualization increases flexibility by allowing creating and retrieving the logical resources based on business requirements.

The Trusted Third Party Monitoring System layer is organized as a tree structure and divided as follows:

- *Infrastructure Monitoring Resources Providers (IMRP)*: represents the centerpiece of our proposed architecture which arranges all the monitoring data and provides procedures for adding and retrieving it. Its interface is the access point for the CSCs and provisions the required data from diverse data storage in a clear way. The setup and monitoring of cloud-based services represent an expansive number of issues as a consequence of various management standards. Numerous specifications have been presented in the most recent decades, however not all of them have been proven to be compelling and any convention can claim to be a standard according to Aceto et al., (2013).

The IMRP is first combined from various sources by the Infrastructure as a Service (IaaS) CSPs required for monitoring data. Information about CSCs and their services provided by IaaS software itself together with information about virtualized infrastructure resources.

- *Cloud Software (S.W)*: provide parameters and supplies CSCs and project details and information about hosts, VMs and images from

VM properties database. Further information about virtualized resources is polled directly from Cloud software that provides more details about the virtual machines and their assigned resources.

- *Virtualization Library*: empowers the CSPs to effectively manage VM templates, vApps, ISO images and scripts with ease. In addition, it stores and manages content from a central location; (VM/Hosts). Also, it deploys VM templates from VM Library directly onto a host or cluster for usage. VMware, (2016).
- *Monitoring Tool*: The data provided by the monitoring tool is retrieved from monitoring plug-in which are deployed on the virtual machines. They monitor services and report the resource utilization of virtual machines at runtime. The plug-ins perform system calls on the (virtual) servers and read out dynamic parameters like the current CPU, memory and I/O utilization of a virtual machine. Whenever a new virtual machine is created in the Cloud S.W has to be registered with the monitoring tool.
- *SLA Manager*: The measurement and management functionalities contributed from the IMRP as shown in Fig.7 includes the metric value for the SLA KPIs; Pricing, Resource Allocator and SLA monitor indicating the SLA parameters that must be guaranteed and representing the multi-attributes to be considered in the Service Negotiation. The metric value calculated through SLA monitor must provide an optimum system performance and can be negotiated during the establishment of a SLA. SLA manager keeps track of SLAs fulfillment between customers and service providers. It also detects the penalty delay and updates the SLA Violation Detection and registers it in the SLA data storage. Furthermore, SLA manager handle the concerns surrounding the SLA Establishment and Service Selection by communicating with the SLA data storage and Monitoring Services.
- *SLA Violation Detection*: related to the monitoring of the allocated services in order to detect, and in some cases avoid, the SLA violation at runtime.
- *Service Selection*: uses an Analytical Hierarchy Process (AHP) selection algorithm that illustrated before in El-Awadi and Abu Rizka (2015) to score services based on the CSCs satisfaction level for each service as a function of service QoS KPIs and the importance of each KPI for the CSCs.
- *SLA Establishment*: we proposed a multi-attributes negotiation mechanism that considers Pricing, Resource Allocator and SLA Monitor Whereas CSPs provide a pre-defined SLA that incorporates fixed price, fixed response time, and some selective performance choices, this may restrict varying service types and expressing

required service level exactly. Thus, SLAs should be adaptable, variable and flexible to personalize service qualities by expenses plans.

6 EXPERIMENT EVALUATION

El-Awadi and Abu-Rizka (2015) had previously employed a benchmark that considers the SAP, SQL, and Oracle as CSCs applications. Such CSCs request infrastructure as an offered service from the CSP with precise specifications. EMC2 Testing Environment Lab had represented the CSP in that study as shown in Fig. 8 designed using the Opnet app.

The supplied infrastructures by EMCs were a CPU, a memory, storage and a network by configuring the host to operate its virtual machine using VMware VSphere.

VMware Vshphere is the leading server virtualization platform enhanced with consistent management that enables the best performance availability and efficiency; resources' utilization and response time, from the infrastructure and applications. (VMware, 2016).

The results of El-Awadi and Abu-Rizka (2015), showed that customers had over-requested resources leading to underutilization of resources and wasting the main advantage of Cloud computing which is pay per use.

Thus, this motivated this research to propose a new framework for monitoring the infrastructure resources provided by the CSPs. Worth mentioning, that the monitoring tool used in that study was also the VMware VSphere which can monitor the performance and optimize the resources' utilization to deliver improved SLA and achieve infrastructure and application availability. (VShpher, VMware, 2016).

The main objective of this framework is first, to help in monitoring the behavior of infrastructure resources for running the database applications; SQL, SAP, and Oracle, second, is to be able to recommend the suitable resource level usage to a

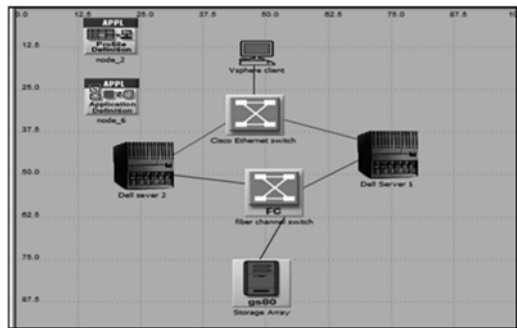


Figure 8. Testing environment.

potential customer requesting the same resources for the same applications.

After monitoring the resources level of utilization for each of CPU, memory and network using VMware VSphere are calculated through Eq.1

$$\text{Resource Utilization} = \frac{\text{Allocated resources}}{\text{pre - defined resources}} \quad (1)$$

The results of this study as presented in figure 9, 10, and 11 respectively shows the allocated, the requested and the recommended usage for each resource.

As seen in Figure 9, the actual CPU utilization exceeded what had been requested by the customer in both SAP and Oracle, therefore according to Cloudorado website, the recommended usage level is restricted to 4GHz for any peak load. (Cloudorado, 2016) but in SQL the recommended CPU need to be decreased to 1 GHz.

The memory utilization as shown in Figure 10 was represented by the consumed and active memory. The active memory is what was actually used by the customers' application while the consumed represent the reserved percentage to the customer application. The results show that there is a large difference between the active and the consumed which indicates that underutilization exists.

Thus the recommended memory usage according to Cloudorado is restricted to all what can be provided by the CSPs denoted by 512GB regardless to the actual customer application level.

As for the network usage given in Figure 11 was signified by the allocated bandwidth values which are restricted by I/O workload of benchmark applications mentioned in El-Awadi and Abu-Rizka (2015).

The values of allocated network bandwidth resulted from the workload that has been generated

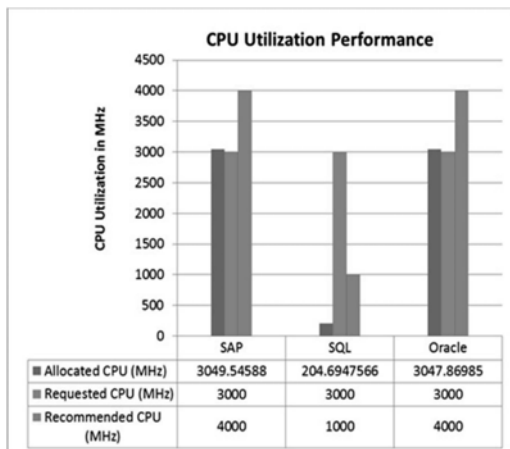


Figure 9. CPU utilization performance.

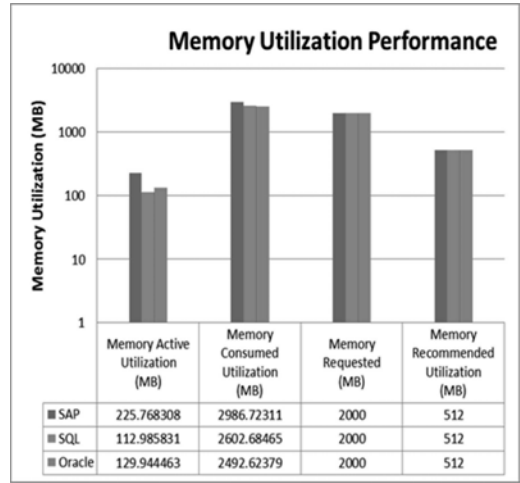


Figure 10. Memory utilization performance.

using IOmeter. On contrast, the values of both requested and recommended network bandwidth are restricted to the minimum of what can be provided by Cloud network providers in Cloud market like offered in VMware network service catalog.

Final, as for the disk utilization that defined by (VMware, 2016) as Disk-I/O counters which support metrics for both physical devices and virtual devices:

A host reads data from a LUN (logical unit number) associated with the physical storage media and a virtual machine reads data from a virtual disk, which is the virtual hardware, presented to the hypervisor running on the virtual machine.

As shown in Figure 12, the amount of allocated disk usage is much less than the required. However, the recommended value is restricted by the minimum amount of disk space provided by the selected CSP (StratoGen, 2016)

This study also contributed with presenting the calculation of VM cost which is divided to acquisition and on-going cost according to (Garg et al., 2011). At the point when assessing on-going costs it is important to look at the route in which the charging will be determined. The premise picked will have a noteworthy impact on how nearly costs track with actual resource utilization, how expectable the invoices will be from month to month, and what endorsement mechanisms will be required to consent variations in consumption.

It is difficult to look at changed costs of services as they offer diverse elements and hence have numerous measurements.

To deal with this challenge, It is defined a volume based metric i.e. cost of one unit of CPU unit, RAM, and network. Therefore, if the VM is priced at p for CPU units, net network units, and RAM memory units, then the cost of VM is calculated in Eq. 2

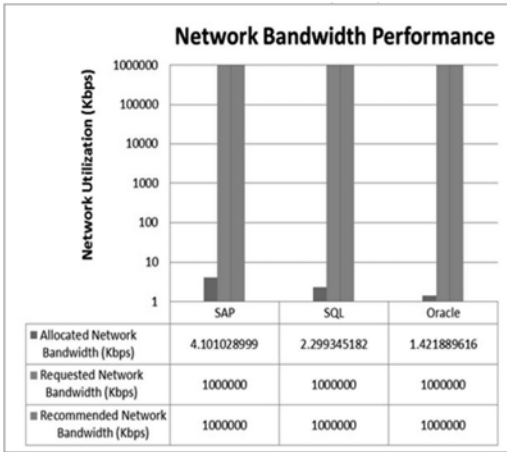


Figure 11. Network utilization performance.

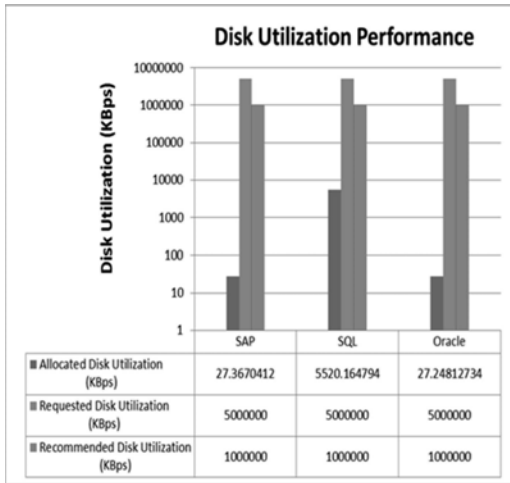


Figure 12. Disk utilization performance.

$$VM \text{ Cost} = \frac{p}{cpu^a * RAM^b * disk^c * net^d} \quad (2)$$

where a,b,c, and d are weights for each resource attribute and $a + b + c + d = 1$. The weight of each attribute can vary from application to application (Garg et al., 2011). For example, for some applications RAM is more important than CPU unit, therefore for them $b > a$. So, we can use different weights of each attribute based on user application. Now, generally CSCs need to transfer data which also incurs cost. Therefore, the total on-going cost can be calculated as the sum of memory, network and compute machine for that particular Cloud provider and service.

The cost for each resource is calculated per-month according to the selected cloud service provider who has been recommended by Clouddorado

recommender website. The selected CSP is StratoGen. StratoGen, (2016) has been defined cost for each resource as shown in Figure 13.

Consistent with CSMIC SMI Measure for Financial Category, (2011) the unit of cost measure is represented in point score ranges from 0 to 10 and Decimal values are not allowed within this range.

Point scores are defined as follows: 0; means no value to the CSC at all and completely insignificant to accomplishing the desired needs and objectives. 2; means minimal value to the CSC compared with other alternatives being considered. 4; means some value to CSC, but slightly less than average compared to other alternatives being considered. 6; means clear value to the CSC, slightly above average compared to other alternatives being considered. 8; means strong value to the CSC, with clear advantages compared to other alternatives being considered. 10; means maximum value to CSC. This ideal score indicates the best fit between the solution being evaluated and the goals of the CSC.

The VM cost of resources for both; the recommended usage as well as the requested are calculated as follows;

$$\text{The requested Cost} = \frac{180}{(96)^{0.6} * (77)^{0.3} * (8)^{0.1} * (0)^0} = 3$$

$$\text{The recommended Cost} = \frac{92}{(72)^{0.6} * (19)^{0.3} * (2)^{0.1} * (0)^0} = 10$$

The point 3 value score proved that minimal value to the CSC compared with other alternatives needs to be enhanced and to be considered rather than the point 10 value score proved that the maximum value to CSC indicates the benefit of the suggested framework for the CSCs furthermore, the importance of the service recommender to let the CSCs get benefits from the pay per use Cloud advantage.

Item	Per Hour	Per Month
CPU (Per vCPU)	\$0.043	\$23.95
RAM (Per GB)	\$0.065	\$38.40
NetApp SAN SAS Disk (Per GB)	\$0.0014	\$0.75
NetApp SAN SATA Disk (Per GB)	\$0.0009	\$0.50
SSD Array Disk (Per GB)	\$0.0028	\$1.49
VMware Enterprise Plus Licensing	INCLUDED	INCLUDED
1Gbps Internet Connection	INCLUDED	INCLUDED
Daily off-site Veeam backup with 7 day retention	INCLUDED	INCLUDED

Per hour pricing is based on metered CPU, RAM and Disk Storage in use by your virtual datacenter.
Per hour plans are paid monthly in arrears (credit card only).

Figure 13. StratoGenIaaS resources prices. [19].

7 CONCLUSION AND FUTURE WORK

The underutilization of resources problem experienced by CSCs resulting from the overestimation of their actual usage of resources, had pushed this research to propose a framework presenting the role of a trusted third party whose responsibility is to monitor the infrastructure resources provided the CSPs. This study also presented the cost equation of calculating the values for both actual and requested costs. The findings of this study indicate that, first a real difference between the requested and the actual usage of resources by the CSCs in both quantity and cost, second, the suggested framework would be of real and significance to the CSCs and a valuable benefit of the role of the trusted third party recommending system. This study highly recommends the implementation of the suggested framework in the Cloud environment while negotiating and establishing the SLA between the CSCs and the CSPs.

In future research, monitoring and evaluating of more than one service KPI is suggested as this study had only focused on the Cloud performance and efficiency measured by the resource utilization and response time.

Also in future research, it is highly recommended to implement the proposed framework on the layer of Software as a Service (SaaS).

REFERENCES

- [1] El-Awadi, R. and Abu-Rizka, M., 2015. A Framework for Negotiating Service Level Agreement of Cloud-based Services. *Procedia Computer Science*, 65, pp. 940–949.
- [2] Habib, S.M., Hauke, S., Ries, S. and Mühlhäuser, M., 2012. Trust as a facilitator in cloud computing: a survey. *Journal of Cloud Computing*, 1(1), pp. 1–18.
- [3] Yaqub, E., Yahyapour, R., Wieder, P., Kotsokalis, C., Lu, K. and Jehangiri, A.I., 2014, June. Optimal negotiation of service level agreements for cloud-based services through autonomous agents. In *Services Computing (SCC)*, 2014 IEEE International Conference on (pp. 59–66). IEEE.
- [4] Messina, F., Pappalardo, G., Santoro, C., Rosaci, D. and Sarné, G.M., 2014, June. An agent based negotiation protocol for cloud service level agreements. In *WETICE Conference (WETICE)*, 2014 IEEE 23rd International (pp. 161–166). IEEE.
- [5] Longo, A., Zappatore, M. and Bochicchio, M.A., 2015, June. Service Level Aware-Contract Management. In *Services Computing (SCC)*, 2015 IEEE International Conference on (pp. 499–506). IEEE.
- [6] Frey, S., Reich, C. and Lühje, C., 2013. Key performance indicators for cloud computing SLAs. In *The Fifth International Conference on Emerging Network Intelligence, EMERGING* (pp. 60–64).
- [6] Son, S. and Jun, S.C., 2013, May. Negotiation-based flexible SLA establishment with SLA-driven resource allocation in cloud computing. In *Cluster, Cloud and Grid Computing (CCGrid)*, 2013 13th IEEE/ACM International Symposium on (pp. 168–171). IEEE.
- [7] Alemeye, F. and Getahun, F., 2015, September. Cloud readiness assessment framework and recommendation system. In *AFRICON*, 2015 (pp. 1–5). IEEE.
- [8] Aceto, G., Botta, A., De Donato, W. and Pescapè, A., 2013. Cloud monitoring: A survey. *Computer Networks*, 57(9), pp. 2093–2115.
- [9] Aceto, G., Botta, A., De Donato, W. and Pescapè, A., 2012. Cloud monitoring: Definitions, issues and future directions. *CLOUDNET*, 12, pp. 63–67.
- [10] Brinkmann, A., Fiehe, C., Litvina, A., Luck, I., Nagel, L., Narayanan, K., Ostermair, F. and Thronicke, W., 2013, December. Scalable monitoring system for clouds. In *Utility and Cloud Computing (UCC)*, 2013 IEEE/ACM 6th International Conference on (pp. 351–356). IEEE.
- [11] Wu, L., Garg, S.K. and Buyya, R., Service Level Agreement (SLA) based SaaS Cloud Management System.
- [12] Montes, J., Sánchez, A., Memishi, B., Pérez, M.S. and Antoniu, G., 2013. GMonE: A complete approach to cloud monitoring. *Future Generation Computer Systems*, 29(8), pp. 2026–2040.
- [13] Server Virtualization with VMware vSphere | United States. 2016. Server Virtualization with VMware vSphere | United States. [ONLINE] Available at: <https://www.vmware.com/products/vsphere>. [Accessed 29 February 2016].
- [14] vSphere Content Library for Content Management: VMware | United States. 2016. vSphere Content Library for Content Management: VMware | United States. [ONLINE] Available at: <https://www.vmware.com/products/vsphere/features/content-library>. [Accessed 29 February 2016].
- [15] Cloud Server Price Comparison | Clouddorado – Find Best Cloud Server from Top Cloud Computing Companies. 2016. Cloud Server Price Comparison | Clouddorado - Find Best Cloud Server from Top Cloud Computing Companies. [ONLINE] Available at: https://www.clouddorado.com/cloud_server_comparison.jsp. [Accessed 29 February 2016].
- [16] Knowledge Base. 2016. . [ONLINE] Available at: https://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=1001805. [Accessed 29 February 2016].
- [17] Garg, S.K., Versteeg, S. and Buyya, R., 2011, December. SMI Cloud: a framework for comparing and ranking cloud services. In *Utility and Cloud Computing (UCC)*, 2011 Fourth IEEE International Conference on (pp. 210–218). IEEE.
- [18] Disk I/O Counters [Storage I/O]. 2016. Disk I/O Counters [Storage I/O]. [ONLINE] Available at: https://www.vmware.com/support/developer/vc-sdk/visdk41pubs/ApiReference/disk_counters.html. [Accessed 11 March 2016].
- [19] VMware Cloud Hosting | Cloud VMWare Hosting | StratoGen. 2016. VMware Cloud Hosting | Cloud VMWare Hosting | StratoGen. [ONLINE] Available at: <http://www.stratogen.com/products/cloud-hosting/vmware-cloud-hosting/>. [Accessed 11 March 2016].
- [20] Selecting a cloud provider. 2016. Selecting a cloud provider. [ONLINE] Available at: <https://slate.adobe.com/a/PN39b/>. [Accessed 11 March 2016].

Human work perspectives in cyber-physical systems impacting industry & business (*with accent on Czech Republic*)

Eva Kasparova

University of Economics in Prague, Prague, Czech Republic

ABSTRACT: The paper is focused on the industrial and business development in the current era called fourth by an industrial revolution. Attention is paid to the impact of technological advances on the nature of human work in the perspectives of the near future. The paper introduces visions of computerization and sensitivity of professions to it. Digitization is the first major challenge but the process is also associated with a number of risks. One of the key problems seems to be the amount of traditional profession reduction, the number of existing profession transformations, jobs cuts and the overall changes in the structure of employment in industrial and business sectors. The findings of a research focused on anticipating a labour market in the Czech Republic are introduced in the paper. “Smart companies” building needs “smarter people”. The very important step to adapt to a digital reality is a sufficient cooperation among industry and business, research centres and educational institutions to build an appropriate and consistent system together. The gap called “valley of death” still exists there. Improving the current situation is not only the task for Czechs but it must be discussed and solved in all countries in our globalised world. The paper opens a space for a discussion of the topic.

1 THE INDUSTRIAL DEVELOPMENT REFLECTION

“Up until a few years ago, the expression “industrial revolution” was associated with history, with the radical societal and economic changes that started some 250 years ago. But suddenly, the term is ubiquitous.” M. Šefčovič (Weforum, ©2016), the vice president of EC, in charge of Energy Union, discusses the current development.

The term industrial revolution is understood as a concept and a development that has fundamentally changed our society and economy. (Bloem et al, 2014) The current development is referred to as the fourth industrial revolution, driven by new technologies.

Klaus Schwab (2016) writes. “Of the many diverse and fascinating challenges we face today, the most intense and important is how to understand and shape the new technology revolution, which entails nothing less than a transformation of humankind. We are at the beginning of a revolution that is fundamentally changing the way we live, work, and relate to one another. In its scale, scope and complexity, what I consider to be the fourth industrial revolution is unlike anything humankind has experienced before.”

Looking back at the development of the industry, the economy and society in the last 250 years mentioned above, history identifies some significant periods of relatively major changes.

In the nineteenth century, the first industrial revolution based on mechanical production equipment driven by water and steam power started and replaced manual labour with industrial methods. At the beginning of the 20th century, the second industrial revolution with the implementation of electrical power developed mass production on assembly lines with strict division of labour. The momentous invention of the transistor in the mid-twentieth century set into motion the third industrial digital revolution with the use of electronics, IT, and robots to further automate production. The current phase of industrial development is based on the use of cyber physical systems (The 4th industrial revolution, ©2013).

The 4th Industrial Revolution is about adopting societal trends in order to change the way products are made. It is a demand-oriented revolution in which the offer is tailored to consumers’ expectations.

S. Gmeiner and G. Frey identify the key characteristics of the current epoch by the following aspects:

- A social production that brings all stakeholders together to boost productivity and competitiveness
- A flexible production that enables to deliver exactly what consumers want and still control costs and make profit

- A smart production that connects products, machines, plants and people
- Producing services that provide consumers with the best experiences and ensure you a better margin
- Leading edge technology again entering a drastic change

It is largely expected and discussed by experts that the physical environment will be integrated with the information network without any problems. The internet is combining processes, systems and intelligent machines to form a sophisticated network. Companies are implementing more and more sophisticated manufacturing software (Industry 4.0, ©2015).

Information and Communication (ICT) Systems can fundamentally transform science, society, economy and all our current institutions. As Dirk Helbing predicts, ICT have the potential to change most of our traditional institutions (Helbing, 2014):

- The way of providing education (personalized education)
- The way of researching (Big Data analytics)
- The way of transportation (Google cars) or transportation of goods (drones),
- The way of shopping (take Amazon or eBay),
- The way of production (3D printers),
- The health system (personalized medicine),
- Politics (citizen engagement)
- The entire economy (with the makers community, the emerging sharing economy, and consumers, i.e. co-producing consumers).
- The way of conducting financial business, which used to be the domain of banks, is increasingly replaced by algorithmic trading, PayPal, Bitcoins, Google wallet.
- The way of conducting insurance business (financial products, credit default swaps).
- The way of waging war is (partially) replaced by cyber war.

Smart devices will take over certain activities previously carried out by people. Methods of machine perception, auto-configuration and diagnostics, and computer-linking of machines and components are envisaged. Machines and made components will have to be able to communicate with each other and the production line will have to be capable of autonomously and dynamically reconfiguring so that it will be possible to produce efficiently and in small series in large plants. Industry will widely use the Internet. Thanks to sensors, cameras, transmitters and reader-codes cyber-physical systems will be racing to a certain extent to manage themselves. Automated warehouses will timely send the order. Components and finished

products will be equipped with microchips and will determine themselves how they should be handled. The machines themselves will log maintainers. Consumer demands will go over the Internet directly to the production line, so that individual orders will be processed at the price of large-scale production (Industrial equipment, ©2002–2016).

2 THE CHANGING ROLE OF THE HUMAN FACTOR

Another group of experts characterizes modern technologies as disruptive technologies. Technologies are seen as a tool which will totally wipe out existing markets and replace them with new, more technologically advanced ones. The impact of this process is already affecting nearly every person. “Disrupt or be disrupted! Technology is your best friend and your worst enemy! Why you need to learn new skill sets to avoid a possible poverty crisis”, warns A. M. Barker (2015).

Therefore the very important result of the digitization is the changing role of the human factor there. Will the transformation to the digital world really be without any serious problems? K. Schwab (2016) writes: “The Fourth Industrial Revolution has the potential to empower individuals and communities, as it creates new opportunities for economic, social, and personal development. But it also could lead to the marginalization of some groups, exacerbate inequality, create new security risks, and undermine human relationships.”

One of the key tasks seems to be the necessity of transforming a wide range of traditional professions and occupations.

A number of unanswered questions must be answered and related problems solved.

- Where will be the current employee replaced by a machine?
- Will the digitization affect employment and employees who are at risk?
- Which professions will be extinguished?
- What qualifications will survive?
- What is the appropriate way to educate employees?
- What types of employees are/or will be needed?
- What role will be played by educational institutions?
- What is the appropriate way of effective cooperation among different sectors?

3 APPROACHES TO COMPUTERIZATION WITH REGARD TO PROFESSIONS

In the next two decades, half of the jobs could be more or less threatened by the increasing influence

of computerization. At least that is a vision based on a study conducted within the Oxford Martin Programme on the Impacts of Future Technology. The authors of the study called “The future of employment: how susceptible are jobs to computerization?” rated 702 occupational groups (derived from the American occupation classification-Standard Occupational Classification (SOC)) by the probability of being affected by computer technology and automation. They identified three barriers that prevent automation of individual work locations.

1. Perception and Manipulation
2. Creative Intelligence
3. Social Intelligence

With the help of this classification Frey and Osborne (2013) estimated the probability of the danger of computerization of each of the 702 occupational groups during the next two decades. As a result they were able to calculate an estimate of the proportion of jobs that can be replaced by a computer or other machines due to the ongoing computerization in the next 20 years. If there are jobs in the afore mentioned technological barriers to high-level computerization this option (to replace humans on the job with computer-controlled machines) is low.

“Study Übertragungsderv on Frey/Osborne (2013) auf Deutschland” which was performed by a major research institution Zentrum für Europäische Wirtschaftsforschung in Mannheim brings different findings. These findings have shown the future in a completely different way. The authors used for the analyses data from international research on adult PIAAC (Programme for International Assessment of Adult Competencies), organized by the OECD.

The authors of the study focused on how often people deal with analytic and interactive tasks and activities in exercising their profession. The presumption is the fact that such tasks are much more difficult to automate. The probability of computerization should not be understood as a probability of certain jobs being replaced by future machines. Rather it is possible to understand it as probability of job computerization as an indicator of identification of concrete workers who are performing tasks or activities that can be potentially automated in the near future. These people will be faced with the challenge of adapting to the technological changes (Bonin, Gregory and Zierahn, 2015).

Another example of an analysis of how digitalization affects the future of jobs inspired by an article by Osborne and Frey (2013) was realised in Norway in 2014.

Norwegian authors identified three interrelated bundles of forces of the digital disruption:

1. exponentially growing mankind’s abilities to produce, store, process, and transmit digitally coded information
2. new important phenomena publicly unknown a decade ago: cloud computing, mobile internet, and social media
3. more intelligent software algorithms; processing capacity and robots becoming mass consumer products

These forces should be seen as enablers; they only have social impact if they are embedded into day-to-day lives of individuals and organizations in such a way that behaviours and structures are adjusted to reflect the possibilities that have opened with technological advances. The Norwegian study follows Osborne and Frey’s (2013) example and method and the approach is applied on the Nordic context. The authors assume a technological capabilities point of view, *i.e.*, they do not consider political or social forces that may influence technology adoption (Pajarinen, Rouvinen and Ekeland, 2014).

4 PROGNOSIS OF COMPUTERIZATION IN THE CZECH REPUBLIC

There is a Czech study of the probability of computerization, undertaken by Fond dalšího vzdělávání MPSV¹ as a part of the PŘEKVAP project (*Předvídaní vývoje trhu práce a zkvalitňování výstupů tohoto předvídaní*. 2015²) assuming that certain professions are more at risk of being computerized than others. The findings of the research are based on a recalculation of the USA results as counted by Frey and Osborne with the help of American O*NET. (Osborne a Frey, 2013). In view of the heterogeneity of forecasts and visions, Czech authors did not concentrate on expressing accurate computerization probability values. Instead they concentrated on individual professional groups and education types in order to identify the most and the least threatened groups based on the observed criteria.

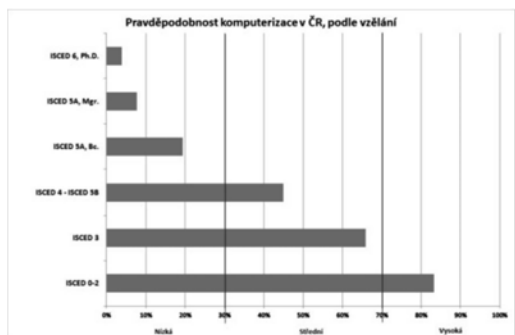
The study measured the relationship between computerization and the level of education in this context.

The results of the research show that the probability of computerization is closely connected to the required education. The below chart indicates that the higher the required education the lesser the probability that human work will be replaced by automated machines.

¹Further education fund of the Ministry of Labor and Social affairs.

²Predicting the development of the job market and improving the outputs of this predicting.

Table 1. Probability of computerization in CR by education.



Resource: FDV MPSV (2015). Předvídání kvalifikačních potřeb.³

There are other questions related to the level of achieved education besides the threat of some professions with a connection to education becoming extinct.

Other important issues that need to be solved include e.g. the potential level of expected computerization or intermediate computerization connected with a certain profession and the required appropriate training for the said profession.

The question is: "Are the "future" experts equipped with sufficient knowledge and skill for working in a digitized and computerized environment?"

Is their training for a profession in a digitized and computerized environment sufficient?

Osborne and Frey (2013) only define the probability of being threatened by computerization for groups of professions. The authors of the Czech study observed the probability of being threatened by computerization for individual branches based on the representation of the groups of professions in any given branch.

The above table depicts 26 branches representing the whole economy of the Czech Republic and the probability of job computerization in each of the sectors.

The findings of the Czech study indicate that in each of the branches, there are certain professions computerizable only with difficulties while on the other hand there are certain easily computerizable professions where computerization can be expected and its probability is very high. It can be seen that high computerization probability is to be found mainly in the branches belonging to the sectors of resources, processing industry and services.

The third table provides an overview of individual professions and the probability of their computerization...

³Predicting qualification needs.

The authors of the Czech study worked with 40 profession groups. The study shows that in the Czech Republic not only professions with lower qualification requirements belonging to ISCO 7, ISCO 8 and ISCO 9 but also professions with higher qualification requirements belonging to ISCO 4 and ISCO 5 find themselves under the threat of computerization. Professions with the highest probability of computerization involve: Administrative workers, secretaries, etc. (ISCO 41), assembly line workers (ISCO 82) and cashiers, ticket booth clerks, etc. (ISCO 52 with the exception of 522).

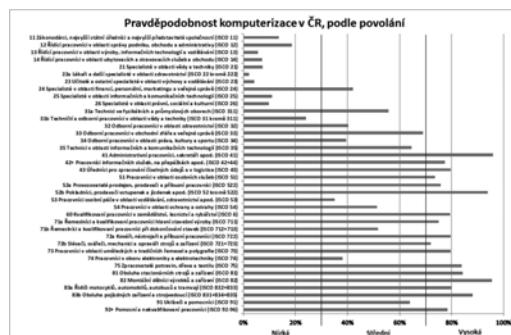
Professions with the lowest probability of computerization include doctors and other experts from the field of health care (ISCO 22 with the exception of 222), Teachers and other experts from the field of education (ISCO 23) and Managerial workers from the fields of production, ICT and education (ISCO 13).

Table 2. Probability of computerization in CR.



Resource: FDV MPSV (2015). Předvídání kvalifikačních potřeb.⁴

Table 3. Probability of computerization by professions in CR.



Resource: FDV MPSV (2015). Předvídání kvalifikačních potřeb.

⁴Predicting qualification needs.

The lowest threat of computerization can therefore be associated with professions requiring mostly university education belonging to the ISCO 1 and ISCO 2. 20. groups (PŘEKVAP, 2015).

These prognoses can certainly be further modified by the rapid development of ICT and quick digitization and computerization of practically all fields of social life. Even now we can estimate that introducing cyber-physical systems into the industry and business will bring higher and higher requirements of professional preparation for adapting to, living in, working with and effectively controlling this environment.

5 CONCLUSIONS

“Smart Industry and Business” will need more “Smarter people”. Investment in education and training seems to be the crucial task of our present global society. These measures and promoting life-long learning can prepare personnel for performing complex work tasks and for working with the new machines. Ultimately, it is the complementary use of labour and machinery, which offers the opportunity for a better job.

The effective implementation of digital technologies:

- Needs a holistic approach in management
- Needs and will do co-operation in the industrial, business, educational, research and governmental sectors
- Needs adequate education and continual training for managers and employees
- Needs effective preparation of students
- Needs centres for innovations
- Needs creation of new options
- Needs an adequate metrics
- Needs tools and platforms
- Needs an effective process of standardization

As Dirk Helbing (2014) writes: “Information systems can help us to manage these challenges, and we can create a lot of jobs with them. We just need to create the right settings.”

REFERENCES

Bloem J., van Doorn M., Duivestijn S., Excoffier D., Maas R. & van Ommeren E. (n.d.), 2014. The Fourth Industrial Revolution [online]. <http://www.fr.sogeti.com/globalassets/global/downloads/reports/vint-research-3-the-fourth-industrial-revolution>.

com/globalassets/global/downloads/reports/vint-research-3-the-fourth-industrial-revolution.

Bonin H., Gregory T., Zierahn U., 2013. Kurzwexpertise Nr. 57, Übertragung der Studie von Frey/Osborne auf Deutschland an das Bundesministerium für Arbeit und Soziales, ZEW, Referat 4, [online]. http://ftp.zew.de/pub/zewdocs/gutachten/Kurzwexpertise_BMAS_ZEW2015.pdf.

FDV MPSV, 2015. Předvídání kvalifikačních potřeb: Koncept – metody – data, Část 3. Čtvrtá průmyslová evoluce a zaměstnanost., [online]. <https://koopolis.cz/sekce/knihovna/407-prekvap-predvidani-vyvoje-trhu-prace-a-zkvalitnovani-vystupu-tohoto-predvidani>.

Frey C. B., Osborne M. A., 2013. The future of employment: how susceptible are jobs to computerisation?, [online]. http://www.oxfordmartin.ox.ac.uk/downloads/acadeaca/The_Future_of_Employment.pdf.

Industry 4.0: The Fourth Industrial Revolution, 2015. MultiCam Canada [online]. <http://multicam.ca/industry-4-0-the-fourth-industrial-revolution/>.

Milton-Barker, Adam, 2015. Disruptive technologies and the 4th industrial revolution are here now, not in the future, how will you prepare yourself, your business and family?, [online]. <https://www.techbubble.info/blog/disruptive-technologies/entry/disruptive-technologies-4th-industrial-revolution-here-now-not-in-the-future>.

Overview, The 4th Industrial Revolution has started, 2002–2016. 3 DassaultSystemes, [online]. <http://www.3ds.com/industries/industrial-equipment/overview/>.

Pajarinen M., Rouvinen P., Ekeland A., 2014. Computerization and the Future of Jobs in Norway., [online]. <http://nettsteder.regjeringen.no/fremtidensskole/files/2014/05/Computerization-and-the-Future-of-Jobs-in-Norway.pdf>.

Schwab K. (n.d.). 2016. The fourth industrial revolution. [online]. http://www.weforum.org/docs/Media/KSC_4IR.pdf.

Schwab, K., 2016. How can we embrace the opportunities of the Fourth Industrial Revolution?, [online]. World Economic Forum, <http://www.weforum.org/agenda/2016/01/how-can-we-embrace-the-opportunities-of-the-fourth-industrial-revolution>.

Šešćović, Maroš, 2016. Will Europe lead the Fourth Industrial Revolution? [online]. World Economic Forum, <http://www.weforum.org/agenda/2016/01/will-europe-lead-the-fourth-industrial-revolution>.

The 4th Industrial Revolution needs POWERLINK and openSAFETY, 2013.06 B&R, [online]. <https://www.brautomation.com/cs/spolecnost/customer-magazine/2013/201306/the-4th-industrial-revolution-needs-powerlink-and-opensafety/>.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Statistical analysis of mobility patterns and passive bandwidth reservation in vehicular networks based on a dynamic programming approach

M. Tropea & F. Strangis

D.I.M.E.S. Department, University of Calabria, Rende, Italy

ABSTRACT: The employment of an appropriate Bandwidth Management Scheme (BMS) is needed in wireless networking, given that the main desire of end-users is to take advantage of satisfactory services, in terms of Quality of Service (QoS), especially when a particular charge is paid to meet the requirement. In this paper we are interested in investigating how the continuity of services can be guaranteed in QoS networks, when users move from a cell to another one, under an infrastructure cellular coverage. The only way to face this issue is represented by the employment of in-advance bandwidth reservations, although it leads the system to waste bandwidth resources, since they are not used until the mobile host enters the coverage cell where the passive request has been made. A new scheme for predicting user movements is proposed, taking the advantage of the dynamic programming approach, that is able to reduce the number of possible roads to be considered and thereby increasing/decreasing the accuracy/redundancy of the proposed model. Several simulation runs have been carried out in order to assess the effectiveness of the proposed idea, having the possibility of optimizing the blocking probability (under 15%), the dropping probability (under 12%) and the average prediction error.

Keywords: vehicular networks, mobility patterns, bandwidth reservations, dynamic programming

1 INTRODUCTION

With the rapid growth of Internet of Things (IoT) and mobile communications, the need for QoS guarantees has become of primary importance, especially when hand-over events occur by Mobile Hosts (MHs) changing coverage areas during their active sessions; they may find scarce resource availability in new locations and the current active connections can be dropped. To the best of our knowledge, the only way to ensure QoS and service continuity to mobile users consists of making a bandwidth reservation over all the cells that a MH will visit during its active connection. There are many protocols able to ensure early reservations like Next Step In Signaling (NSIS) [1], Dynamic ReSerVation Protocol (DSRVP) [2] and Mobile ReSerVation Protocol (MRSVP) [3], but a prediction scheme is mandatory in order to know which coverage cells a user will probably visit during its Call Life Time (CLT). On the basis of previous works [4], [5], we considered the MRSVP, which gives the possibility to exchange the right communication messages among the predicted coverage cells, achieving the needed passive amount of bandwidth in the cells where the MH will probably hand-in. The same Markov model has been

considered, but an optimization on the number of chain states is now proposed: in the previous contributions, only one hand-over direction has been considered for the hand-off event toward a next cell, without considering the roads topology that characterize MH movements. Given that the number of chain states could be very large if all the roads that lead to another cell are considered, an optimization scheme is proposed. In particular, the dynamic programming approach is considered [6], having the possibility to choose the right number of states for the Markov model, taking into account the morphology of the considered geographical region. An approximation has been introduced and the associated error has been minimized. Clearly, in order to implement and realize this kind of prediction, a real network operator has to analyse users' mobility, through a statistical treatment. In our case, without access to real data about MH movements, we employed the Citymob for Roadmaps (C4R) mobility generator [7], in order to appreciate prediction performance when mobility traces are extracted from real roadmaps of different countries (the mobility model has a heavy impact on the obtained results, that may be unsuitable if the adopted mobility model is unrealistic). The integration between the Markov process

and the dynamic programming approach leads to a new distributed prediction scheme, called Dynamic Markov Prediction Algorithm (DMPA), tested through extensive simulation studies. The rest of the paper is organized as follows: section II gives an overview of the existing related work, section III gives a detailed description of the proposed scheme, by considering the environment and the solution. Section IV shows our simulation results, then section V concludes the paper.

2 RELATED WORK

Mobility and resource management are critical for providing QoS guarantees in wireless networks, so it is very important to accurately describe mobility patterns of MHs in wireless cells, especially when a prediction approach is needed. In [8] the Mobility-Dependent Predictive Resource Reservation (MDPRR) scheme is proposed, that is able to provide flexible usage of limited resource in mobile multimedia wireless networks. Each cell is divided into non-hand-off, pre-hand-off and hand-off zones, so that bandwidth is reserved in the target/sub-target cell as mobile stations move into the pre-hand-off zone. An admission control scheme is also considered to further guarantee the QoS of real-time traffic as, for example, Voice over IP, as proposed in [9] and [10]. The Fixed Bandwidth Reservation (FBR) scheme [11] can improve the dropping probability of hand-off connections by reserving a fixed number of channels exclusively for hand-off connections. The drawback of this scheme is that the reserved bandwidth is often wasted in the hot spot area. In [12] the authors optimize some system parameters in terms of Call Dropping Probabilities (CDPs) and Call Blocking Probabilities (CBPs) introducing a prediction algorithm based on data mining approaches, in order to implement a distributed Call Admission Control (CAC) scheme, considering also the throttle flag as indication of the usage of each cell. Through estimation of MHs trajectory and arrival/departure times in [13], a group of future cells is determined: it constitutes the most likely cluster into which a terminal will move. Two passive reservation techniques are proposed in [14], exploiting Wiener prediction and time series theory, making in-advance reservations under non-Poisson and/or non stationary arrival processes, arbitrary distributed call and channel holding time and arbitrary per-call resource demands. In [15] the authors give a contribution in WLAN infrastructure planning, basing their decisions on mobility prediction: they propose a new method for feature extraction with a novel neural network classifier based on a hidden genetic algorithm, reaching an

acceptable prediction accuracy. In previous works, like [16] and [17], a prediction technique based on the Cell Stay Time (CST) evaluation of a mobile user is proposed. A formula that relates cell coverage radius and speed is calculated and resource reservation techniques have been proposed, so it is possible to evaluate the number of coverage cells that users will visit during their CLT. To the best of our knowledge, all the literature is focused on the prediction of a single next cell, without the guarantee of service continuity during the whole flow lifetime. In addition, they do not take into account the geographical morphology of the considered region, in terms of roads, that heavily influences driving styles and mobility patterns in terms of cell sequences. In this work, instead, the DMPA algorithm is proposed: it provides a distributed set of Markovian predictors, each one optimized in terms of number of states and local road topology coverage. With respect to previous works [4], [5], [17], [18], DMPA optimizes the number of states for each chain, taking into account the particular roads structure. In this way, the number of states for each cell is variable and it strictly depends on the possible MH movements in the considered region. As mentioned before, the number of states of each chain is adequately chosen, minimizing the error committed during the approximation. Each Markovian chain is trained by taking into account local trajectories (belonging to the associated coverage cell); each predictor is specialized for the specific coverage area, with different traffic densities, in terms of roads, road populations, moving directions and so on; the considered signaling protocol has been integrated with Markov chains in order to realize a complete prediction scheme. It must be outlined that the proposed scheme (a prediction scheme in general) is very suitable for enhancing vehicular applications [19–22], given that the a-priori knowledge could lead to an early optimization of a particular application [23]. Although the proposed idea is based on MRSVP and Markovian processes, it is suitable for any other signaling protocol and/or (un)conventional prediction approach. The effectiveness of DMPA has been also verified in terms of accuracy error, by considering different movement traces of MHs and the length of learning observations.

3 PROBLEM STATEMENT, SYSTEM CHARACTERIZATION AND PROPOSED IDEA

In this section, the proposed idea is completely described. It must be noted that the proposed idea does not depend on the employed protocol: for example, it can be one of those used or described in

[1], [3]. As stated before, we chose the MRSVP [1], with which one reservation is made by a user on the current coverage cell (active reservation), while passive ones are made on the predicted remote cells. When hand-overs events have to be managed in an adequate manner, MRSVP can be employed, handling users mobility and offering guaranteed services, giving the chance to mobile users to make reservation requests over more than one cell, by their proxy agents. For more details about MRSVP to see [3]. In our work, we considered that a MRSVP session starts with the active service request performed by a MH u on its active cell c_i ; if there are no free channels in c_i , the call is refused, else c_i applies the results obtained in [3], [16] to evaluate the number of predicted hand-over events. If no hand-over events are predicted (the $CST \gg CLT$), then the call is accepted (u will visit only the current cell c_i). Otherwise the proposed DMPA is used to predict the neighbor cell $nc \in Adj(c_i)$, where $Adj(c_i)$ is the set of neighbors of cell $c_i \in C$ and $\|Adj(c_i)\| = n$, where n is the number of possible hand-over directions. We considered a generic Geographical Region GR covered by a number of cells equals to c . Let C be the set of coverage cells of the considered wireless network, $C = \{c_1, c_2, \dots, c_c\}$ with $\|C\| = c$. For each cell $c_i \in C$, with a coverage radius r , a set of neighboring cells $Adj(c_i)$ can be defined, on the basis of network topology and cell adjacencies. A circular coverage cell can be approached with a n -edge regular polygon and, considering $n = 6$, coverage cells are represented by regular hexagonal areas, as approached in [18]. In addition, a set S_{ho} of n movement directions $d_1 \dots d_n$ can be introduced, where $d_j = \theta \cdot (2 \cdot j - 1) / 2 \text{ rad.}$, $\theta = 2\pi/n \text{ rad.}$ and $j = 1..n$ (it represents the j -th side of the hexagon), so $S_{ho} = \{d_1, \dots, d_n\}$, $\|S_{ho}\| = n$. In this work $\|Adj(c_i)\| = \|S_{ho}\| = 6, \forall c_i \in C$.

In previous works [4], [5], [17], [18] we did not differentiate the model for different road densities. In particular, referring to Fig. 2, we can observe how a cell $c_i \in C$, based on the value of r , can manage a different number of roads. Let us consider two cells $c_1, c_2 \in C$, which cover real geographical areas (two locations of a city in south Italy are

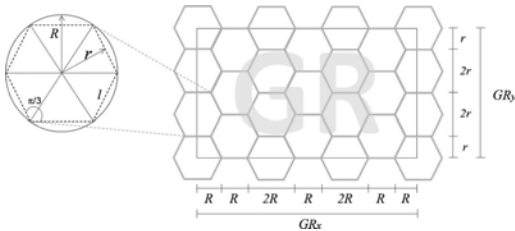


Figure 1. Hexagonal approximation ($n = 6$) and GR coverage.

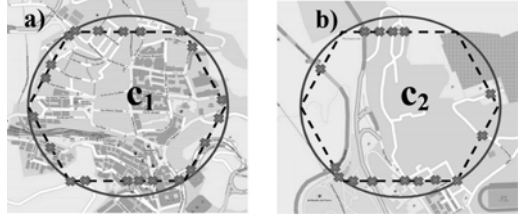


Figure 2. Different road densities for different coverage cells c_1 and c_2 .

considered). We can immediately observe how the number of possible hand-over roads for c_2 (three sides on a total of six have only one possible direction for hand-over) is less than the one for c_1 (the crosses on the hexagonal sides represent some of the possible hand-in/hand-out points, i.e. intersections among roads and cell sides).

The main idea is to extend the number of states of the model in order to take into account all the possible crossing directions; on the other side, the complexity of the proposed model cannot be increased indefinitely, so a right trade-off should be found, aggregating, when possible, roads information belonging to users mobility. At this aim, we considered the approach of [6], in which an input sequence of a certain size has been divided into a lower number of sub-sequences, each one represented by the average value; the obtained partitioning minimizes the error due the approximation process. Let us hypothesize that each coverage node (Access Point, Base Station, etc.) is able to recognize the direction on which a MH enters or leaves the cell (many Direction-of-Arrival (DoA) algorithms are present in the literature, depending on the adopted technology). So, referring to a generic coverage cell $c_i \in C$, for each $d_j \in S_{ho}$ we can define a set of roads $RD_{dj} = \{rd_{dj1}, rd_{dj2}, \dots, rd_{djJ}\}$ where $rd_{dj1} \in [0, 2\pi]$, $k = 1, \dots, dj_j$.

From Fig. 3 it can be seen how, for $n=6$, for each $d_j \in S_{ho}$ which represents the “average” direction associated to side j , the lower and upper bounds can be determined as $[d_j - \pi/6, d_j + \pi/6)$, so each $rd_{dj1} \in RD_{dj}$ belongs to that interval. Figure 3 shows, on the right, how the angles of road intersections can be determined. Given the sequence of roads/angles $RD_{dj} = \{rd_{dj1}, rd_{dj2}, \dots, rd_{djJ}\}$, with $\|RD_{dj}\| = J$ and a compression factor λ (with $\lambda < J$), the set RD_{dj} has to be divided into λ sub-sequences and each of them has to be replaced with its average value. We followed the approach of [6], which is able to solve a subclass of the run-length coding scheme in polynomial time, using a dynamic programming approach. In particular, each $rd_{dj1} \in RD_{dj}$ is associated to the terminal nodes of the base level of a Compact Binary Tree (CBT), composed

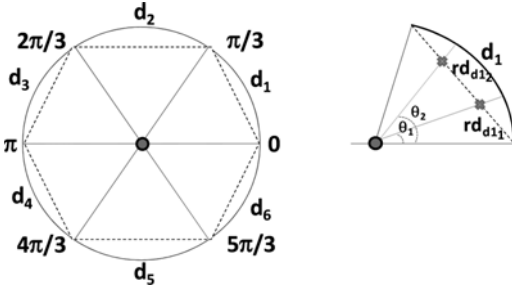


Figure 3. Cell directions subdivision and intersection degrees determination.

by $2 \cdot J - 1$ nodes, assuming that $J = 2^{h_j}$. The CBT is represented by an array of the form $T_{d_j} = [t_1, t_2, \dots, t_{2^{J-1}}]$, where each element is associated to a node: the last J elements store the values of RD_{d_j} , while each t_k , with $k < J$, has two children t_{2k} and t_{2k+1} and 2^{h_k} descendants, with $h_k = J - \lfloor \log_2 k \rfloor$, which represents a subsequence S_k of the input sequence:

$$S_k = \{t_h \mid t_h = rd_{d_{j_h-j}} \text{ and } k \cdot 2^{h_k} \leq h < (k+1) \cdot 2^{h_k}\} \quad (1)$$

In addition $t_k = (t_{2k} + t_{2k+1})/2 = \mu_k$. Considering c_1 and c_2 and directions d_1 and d_6 respectively, Fig. 4 shows the CBTs. The related arrays are: $RD_{d_1} = \{rd_{d_{11}}, rd_{d_{12}}\} = \{0.1285, 0.8213\}$, $T_{d_1} = [0.4749, 0.1285, 0.8213]$, $RD_{d_6} = \{rd_{d_{61}}, rd_{d_{62}}, rd_{d_{63}}, rd_{d_{64}}, rd_{d_{65}}, rd_{d_{66}}\} = \{1.5012, 1.6081, 1.712, 1.8373\}$, $T_{d_6} = [1.664651, 1.55465, 1.77465, 1.5012, 1.6081, 1.712, 1.8373]$.

With the approach of [24], the problem is solved by minimizing the quantity $ERR(k, \lambda)$, representing the error of compressing the roads subsequence S_k using λ values:

$$ERR(k, \lambda) = \begin{cases} \varepsilon_k & \lambda = 1 \\ \min_{1 \leq p < \lambda} [ERR(2k, p) + ERR(2k+1, \lambda - p)] & \lambda > 1 \\ 0 & \lambda > 2^{h_k} \end{cases} \quad (2)$$

where ε_k is the mean square error committed with the compression of the roads subsequence S_k with a single value:

$$\varepsilon_k = \frac{1}{\|S_k\|} \sum_{t_h \in S_k} (\mu_k - rd_{d_{j_h}})^2 \quad (3)$$

For more details about dynamic programming and run-length coding approach, please refer to [6].

At this point, for each cell $c_i \in C$, an array $\Lambda_i = [\lambda'_1, \dots, \lambda'_n]$ (with $n=6$ in our case) can be defined, where each λ'_j indicates the best compression factor

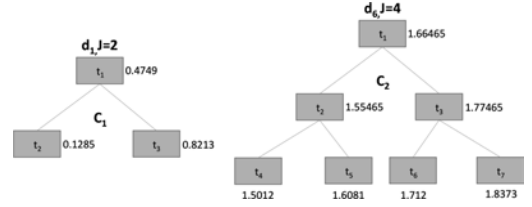


Figure 4. CBTs for two different directions.

for cell c_i associated to RD_{d_j} on direction $d_j \in S_{ho}$ and $1 \leq \lambda'_j \leq \|RD_{d_j}\|$. For each λ'_j , a partition vector $M'_j = [\mu'_{j1}, \dots, \mu'_{j\lambda'_j}]$ represents the compressed sequence, for the j -th side of cell c_i ; each element μ'_{jk} has an associated partition range $p'_{\mu'_{jk}}$, belonging to the j -th Partition Set PS'_j , defined as:

$$p'_{\mu'_{jk}} = \begin{cases} [d_j - \frac{\pi}{6}, \mu'_{jk} + \frac{(\mu'_{j_{k+1}} - \mu'_{j_k})}{2}] & k = 1 \\ [\mu'_{j_k} - \frac{(\mu'_{j_k} - \mu'_{j_{k-1}})}{2}, \mu'_{j_k} + \frac{(\mu'_{j_{k+1}} - \mu'_{j_k})}{2}] & 1 < k < \lambda'_j \\ [\mu'_{j_k} - \frac{(\mu'_{j_k} - \mu'_{j_{k-1}})}{2}, d_j - \frac{\pi}{6}] & k = \lambda'_j \end{cases} \quad (4)$$

In DMPA, a Finite State Markov Chain is considered: the set of states is not only related to the number of possible hand-over directions, but it considers also the number of roads of the covered region. Given vectors M' and PS' , defined in previous section, with $\|PS'\| = \|M'\| = n$, the idea is to associate one state of the Markovian model to each partition subset, representing a compressed set of roads, for each side of c_i . A Markov chain (Fig. 5) can be associated to a cell c_i . A road $rd_{d_{jq}}$, that intercepts cell c_i on side j , is said to belong to state s_{jk} if:

$$rd_{d_{jq}} \in p'_{\mu'_{jk}} \quad (5)$$

In this paper we are not focusing on the definition of a Markovian model, we want to optimize, instead, the number of states of the model. So, without entering in the particulars of the Markovian theory, we can write that the DMPA Markov Chain, related to the t -th (indicated with DMCT), can be described by three terms: Π_t , σ_t and ST_t . For the details about the introduced triplet and their evaluation, please refer to [5]. Figure 6 illustrates how a distributed set of MCs $MC = \{MC_t, 1 \leq t \leq c\}$ can be used to model the whole cellular system. In order to be admitted into the system, each mobile host makes a reservation request to the current coverage cell (active reservation) and to the predicted ones (passive reservations). This

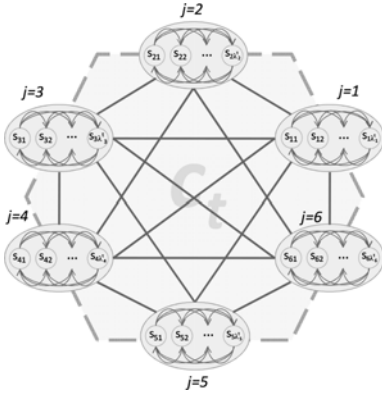


Figure 5. An example of Markovian cell modeling.

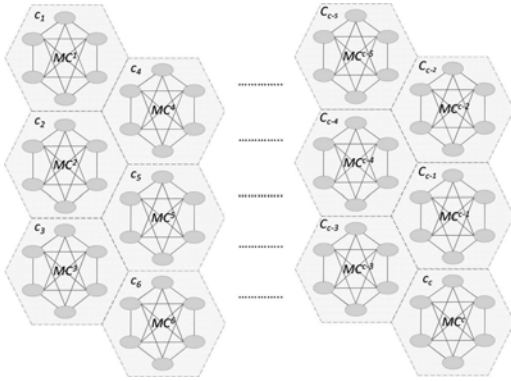


Figure 6. An example of wireless cellular system modeled through MCs.

is made by employing the native signaling packets of the MRSVP. If at least one cell sends a negative answer (no available bandwidth), the call is refused, then the MH will try again later. In the next section, more details about our simulation setups and results will be given.

4 PERFORMANCE EVALUATION

In order to evaluate the proposed integration in terms of average prediction error, Call Dropping Probability (CDP) and Call Blocking Probability (CBP), we considered real mobile environments: Citymob mobility generator [7] and the C4R GUI have been considered, because they give the opportunity of obtaining mobility traces from real maps. In particular, we used maps of some European cities (about 1 km² for each scenario), over which a set of coverage cells (all with the same coverage

radius R) has been considered ($r_i=R, \forall c_i \in C$) and $R \in [50, 250]$ meters. Square maps have been. It is shown that, when the coverage is set-up, the effective number of employed cells always respects the obtained bounds. As stated before, different cities have been considered and, for all of them, obtained results are comparable: without loss of generality, we show the obtained curves for the city of London; Fig. 7 illustrates the obtained coverage for $R=110$ m and $GR_x = GR_y \cong 1000$ m; in this case $c_{low} = 31$, $c_{high} = 54$ and $c = 42$. Once the topology of GR has been determined, as well as the coverage map, the compression algorithm needs to be executed for all cells. In order to choose the right number of partitions for each c_i and direction, a compression factor cf_i is chosen, so:

$$\lambda_i^j = \begin{cases} 0 & \text{if } \|\text{RD}_{d_j}\| = 0 \\ 2^{\lfloor \log_2 \left[(1-cf_i)^{\|\text{RD}_{d_j}\|} \right] \rfloor} & \text{if } \|\text{RD}_{d_j}\| \neq 0 \text{ and } (1-cf_i)^{\|\text{RD}_{d_j}\|} > 1 \\ 1 & \text{else} \end{cases} \quad (6)$$

where the $\lfloor \cdot \rfloor$ operator indicates the integer part.

For each cell c_i , a total of twenty slots ($ns^t = 20$) has been considered and each reservation occupies a single slot in each cell (active or passive). Only for example, Table 1 resumes the values of $\|\text{RD}_{d_j}\|$ and λ_i^j for three cells (c_{10} , c_{18} , c_{33}) as illustrated in Fig. 8, while Table 2 indicates the obtained \mathcal{A}_i sets for different values of cf_i .

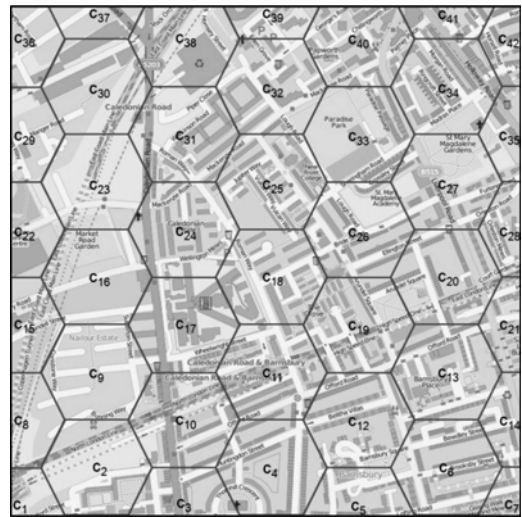


Figure 7. The set of 42 cells used to cover the considered region.

Table 1. Number of roads for each direction for cells c_{10} , c_{18} and c_{33} .

c_i	d_1	d_2	d_3	d_4	d_5	d_6	$\ RD_{d_i}\ _{10}$	$\ RD_{d_i}\ _{18}$	$\ RD_{d_i}\ _{33}$
10	4	4	2	3	3	3			
18	2	1	3	0	1	2			
33	1	2	1	1	3	2			

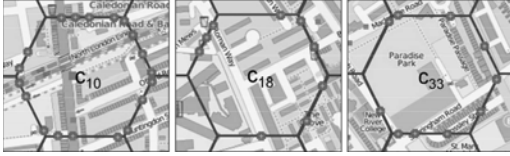


Figure 8. Road sets for some cells of the considered network.

Table 2. The number of compressed roads set for the considered cells.

c_i	$cf_i=0,2$	$cf_i=0,4$	$cf_i=0,6$	$cf_i=0,8$	A_{10}	A_{18}	A_{33}
10	{4,4,2,2,2,2}	{2,2,2,2,2,2}	{2,2,1,2,2,2}	{1,1,1,1,1,1}			
18	{2,1,2,0,1,2}	{2,1,2,0,1,2}	{1,1,2,0,1,1}	{1,1,1,0,1,1}			
33	{1,2,1,1,2,2}	{1,2,1,1,2,2}	{1,1,1,1,2,1}	{1,1,1,1,1,1}			

In order to better understand how the different road sets are partitioned, cell c_{10} has been considered graphically, Fig. 10, for different compression factor values (0.2, 0.4, 0.6 and 0.8). First of all, a training campaign was performed in order to obtain the elements of DMC_i and ST_i in function of the considered map, with $c = 42$.

The Kraub mobility model has been considered [7] (with Acceleration = 1.4 m/s², Deceleration 2 m/s², $s_i = 0.5$ and $\tau = 0.2$ s) for 1000 simulations with 2000s of duration and 270 vehicles for each run (low dense roads have been considered for this simulation campaign). Mobility log files have been obtained and, then, the coverage set of cells has been considered. Different compression factor cf_i values have been used, as well as different coverage radius values R . We assume that each cell is able to recognize the possible roads with DoA or Angle-of-Arrival (AoA) approaches.

Fig. 9 shows how the system responds to DMPA in terms of system utilization, calculated as the average of the ratio between the active bandwidth slots and the total ones, for each cell. It is evident how, in general, the system is under-utilized ($u_{\%}$ belongs to the range [25, 75]). There is an increasing trend for higher coverage radius: when the number of cells decreases from 42 to 10, there is a lower number of passive reservations (lower cells have to be predicted because of the higher geographical covered region). In this case, also the protocol overhead decreases, because there is a lower number of cells among which the signaling

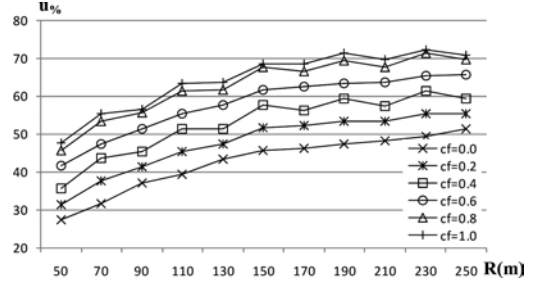


Figure 9. Average system utilization for different values of R and compression factor cf .

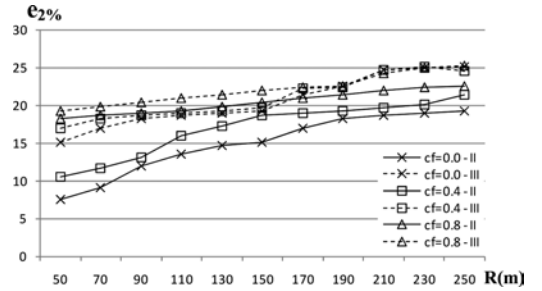


Figure 10. Average prediction error for 2nd and 3rd hand-over events with different values of R and compression factor cf .

packets have to be exchanged. For the same reasons, $u_{\%}$ increases when the algorithm employs a higher grade of roads compression: when $cf \rightarrow 1$, the number of partition sets for each side $\|RD_{d_i}\| \rightarrow 1$, so only one possible direction needs to be considered for each side and the overhead is reduced. Figure 10 gives a description of the trend of the prediction error for the second ($e_{2\%}$) and third ($e_{3\%}$) hand-over events, given that $e_{1\%} = 0$. For a single simulation, it is evaluated as the ratio among the number of users that do not find a passive reservation after the hand-over event and the number of total 2nd or 3rd hand-over events during simulation time. The trend is increasing both for higher coverage radius (host movements are more casual if the considered area is larger) and cf values (system loses the granularity about users movements). The maximum obtained value is 25.3%.

In Fig. 11 the course of the CBP is illustrated. It is evident how for larger R it decreases, because each cell will cover a larger geographical area (so it will serve more users), while the number of available slots remains the same. In addition, for higher cf values, there is a slight decreasing of CBP because, for higher prediction values, the system

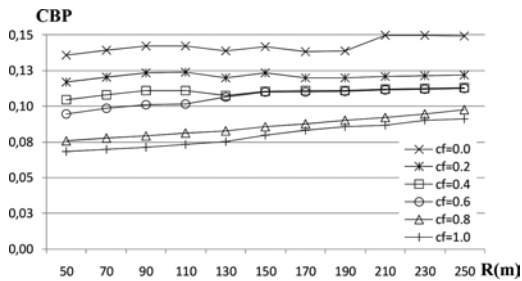


Figure 11. Trend of CBP vs R and cf .

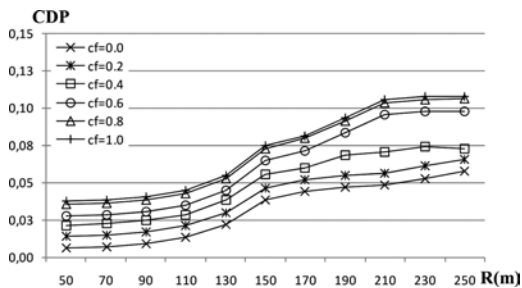


Figure 12. Trend of CDP vs R and cf .

overestimates the available resources and admits more users. Figure 12 shows the trend of the CDP for different values of R and cf . For small coverage areas there is a negligible probability of call dropping (below 5%) because of (based also on the trend of prediction error) more deterministic host movements in the network. In addition, a higher value of cf brings the prediction algorithm to lose more roads information granularity, arriving to the simplest case of one direction for each coverage side.

5 CONCLUSIONS

This work proposes a new Markovian prediction model, DMPA, optimized in terms of number of states. The dynamic programming approach is employed to compute an adequate number of states, able to reflect roads topology properties and mobile hosts behavior. It is also able to guarantee service continuity in QoS networks, without disrupting system utilization performance. The main strength of DMPA resides in the integration of the Markov predictor and the time roads compression scheme, leading to very good performance in terms of prediction error, utilization, CBP and CDP. After many considerations regarding DMPA performance, we highlighted that that a compression

factor of 0.4 can be enough to reduce the number of computations of the Markov model, ensuring a good trade-off in performance results. Future efforts will be focused on the analysis of the proposed scheme, with the main aim of better optimizing the introduced scheme. The dynamic programming approach will be enhanced in order to better consider how the density of roads is distributed along each coverage side.

REFERENCES

- [1] X. Fu, S. Henning, Bader A., Hogrefe D. (2005, Oct.). NSIS: a new extensible IP signaling protocol suite. IEEE Comm. Magazine. 43(10), pp. 133–141.
- [2] Q. Huang, G.S. Kuo, “Dynamic RSVP extension for wireless mobile IP networks,” in Proc. IEEE VTC, 2004, Vol.4, pp. 2683–2687.
- [3] Fazio, P., Marano, S., “A new Markov-based mobility prediction scheme for wireless networks with mobile hosts”, (2012). Proceedings of the 2012 International Symposium on Performance Evaluation of Computer and Telecommunication Systems, SPECTS’12—Part of SummerSim 2012 Multi conference, art. no. 6267025.
- [4] F. De Rango, P. Fazio, S. Marano, “Cell Stay Time Prediction for Mobility Independent Predictive Services in Wireless Networks,” IEEE Wireless Communications and Networking Conference (WCNC2005), New Orleans, Los Angeles, USA, 13–17 March 2005.
- [5] Fazio, P., De Rango, F., Selvaggi, I., “A novel passive bandwidth reservation algorithm based on neural networks path prediction in wireless environments”, (2010) Proceedings of the 2010 International Symposium on Performance Evaluation of Computer and Telecommunication Systems, SPECTS’2010, art. no. 5588656, pp. 38–43.
- [6] G. Shivaram, G. Seetharaman, T.R.N. Rao, “Data compression of discrete sequence: a tree based approach using dynamic programming”, SIAM—ACM, 8th Symposium on Discrete Algorithms, 1997.
- [7] Martinez, F.J., Cano, J.-C., Calafate, C.T., Manzoni, P., “CityMob: A Mobility Model Pattern Generator for VANETs”, ICC Workshops 2008, pp. 370–374, Beijing.
- [8] L. Lu, J. Wu, W. Chen, “The study of handoff prediction schemes for resource reservation in mobile multimedia wireless networks,” International Journal of Communication Systems, vol. 17, pp. 535–552, 2004.
- [9] Hegr, T., Bohac, L., Kocur, Z., Voznak, M. and Chlumsky, P., “Methodology of the direct measurement of the switching latency,” Przeglad Elektrotechniczny, 89(7), 59–63 (2013).
- [10] Frnda, J., Voznak, M., Fazio, P. and Rozhon, J., “Network performance QoS estimation,” 2015 38th International Conference on Telecommunications and Signal Processing, TSP 2015, art. no. 7296443.
- [11] B. Epstein, M. Schwartz, “Reservation strategies for multi-media traffic in a wireless environment,”

- Proceedings of the 45th IEEE VTC, Chicago, U.S.A., pp. 165–169, 1995.
- [12] Chen-Feng Wu, Liang-T. Lee, Hung-Y. Chang, and Der-F. Tao “A Novel Call Admission Control Policy Using Mobility Prediction and Throttle Mechanism for Supporting QoS in Wireless Cellular Networks”, *Journal of Control Science and Engineering*, Volume 2011, 11 pages.
- [13] A. Aljadhai, T. Znati, “Predictive mobility support for QoS provisioning in mobile wireless environments,” *IEEE JSAC*, vol. 19, pp. 1915–1931, 2001.
- [14] T. Zhang et al., “Local Predictive Resource Reservation for Handoff in Multimedia Wireless IP Networks,” *IEEE Journal on Selected Area in Communications*, vol.19, no.10, Oct.2001, pp. 1931–1941.
- [15] Velmurugan, L. and P. Thangaraj, “A Hidden Genetic Layer Based Neural Network for Mobility Prediction”, *American Journal of Applied Sciences* 9 (4): 526–530, © 2012 Science Publications.
- [16] F. De Rango, P. Fazio, S. Marano, “Cell Stay Time Analysis under Random Way Point Mobility Model in WLAN Networks”, *IEEE Communication Letters*, Vol.10, Issue 11, pp. 763–765, Nov. 2006.
- [17] F. De Rango, P. Fazio, S. Marano, “Mobility Prediction and Resource Reservation in WLAN Networks under a 2D Mobility Models,” 63rd Vehicular Technology Conference (VTC Fall), Canada, Sept. 25–28, 2006.
- [18] P. Fazio, F. De Rango, S. Marano, “2D Movement Direction-Based Reservation Scheme for WLAN Clusters with Passive Advanced Reservations,” in *IEEE Canadian Conference on Electrical and Computer Engineering (CCECE 2008)*, Niagara Falls, Canada, May 4–7, 2008.
- [19] A.F. Santamaria, C. Sottile, F. De Rango, S. Marano, “Safety Enhancement and Carbon Dioxide (CO₂) reduction in VANETs,” in *Mobile Networks and Applications (MONET)*, Vol.20, Issue 2, 1 Apr. 2015, pp. 220–238.
- [20] P. Fazio, F. De Rango, A. Lupia, “Vehicular networks and road safety: An application for emergency/danger situations management using the WAVE/802.11p standard,” in *Advances in Electrical and Electronic Engineering*, Vol. 11, Issue 5, 2013, pp. 357–364.
- [21] P. Fazio, F. De Rango, C. Sottile, “An on demand interference aware routing protocol for VANETS,” in *Journal of Networks (JNW)*, Vol. 7, Issue 11, Nov. 2012, pp. 1728–1738.
- [22] F. De Rango, F. Veltri, P. Fazio, S. Marano, Two-level trajectory-based routing protocol for vehicular ad hoc networks in freeway and Manhattan environments, in *Journal of Networks (JNW)*, Vol. 4, Issue 9, 2009, Pages 866–880.
- [23] M.T. Alrefaie, I. Carreras, F. Cartolano, R. Di Cello, F. De Rango, “Map matching accuracy: Energy efficient location sampling using smartphones,” in 6th Int. IEEE Conference on Intelligent Transportation Systems: Intelligent Transportation Systems for All Modes, ITSC 2013; The Hague; Netherlands; 6–9 Oct. 2013.

WARM in the city: WAsTe Route Management in the smart city (WARM City)

M. Tropea, A.F. Santamaria & S. Marano

D.I.M.E.S. Department, University of Calabria, Rende, Italy

ABSTRACT: This extended abstract faces with the problematic of waste management optimization in order to better manage the transportation of waste trucks and then the organization of truck fleet. The main idea is to make intelligent the garbage collectors, equipping them with wireless sensor devices able to create a mesh network in line with the concept of machine-to-machine applications and able, also, to communicate with a central coordinator or Head Quarter (HQ). Thanks to this capability, the garbage collectors can create a wireless network for collecting a set of data used for better performing waste trucks tour. The costs preview for this proposal concern with the purchase of wireless device. In order to perform the opportune choice, it is necessary to realize a software simulator for being able to perform the correct choice on software basis. The importance of making intelligent the garbage collectors is due to the ability of optimizing waste collectors route in order to improve the air quality of a city through a better use of waste trucks and truck fleet that thanks to an optimized path and an adequate number of vehicles are able to reduce pollution emissions.

1 INTRODUCTION

Waste production has been increasing in the last few years, along with economic growth. This has led to the need and subsequent development of efficient waste management solutions. Waste management involves not only the collection, but also the transportation, recycling and disposal of generated waste. The two main issues regard the waste collection and the waste transportation in order to face with the introduction of the new concept of smart cities. Waste collection regards the possibility by operators of controlling the rubbish levels in order to track the container in the truck collection tour. In order to improve the waste management processes, the optimization of routes system is an important field of study. It is important to implement a software platform that permits the control of the containers waste level. All this is possible through the utilization of wireless devices for each container that allow of gathering information in order to develop a container map for optimizing the trash collection routes. The gathered information gives the possibility of performing a series of studies in order to improve the waste management processes exploiting the new concept of Machine-to-Machine (M2M) communications that involves the so-called Internet of Things (IoT): it regards the basic idea of the pervasive presence around us of a variety of things or objects which are able to interact with each other and cooperate with their

neighbors to reach common goals. This permits of improving the quality of life of citizens and make the container equipped with an on-board intelligence, that can be exploited through a series of mechanisms. The optimized management of trash collection route might reduce even further the collection costs and also can improve the air quality of a city through a better use of waste trucks and truck fleet that thanks to an optimized path and an adequate number of vehicles are able to reduce pollution emissions. The optimizing waste collection routes issue involves deciding, for example, which streets must be followed by each garbage truck, which containers should be collected by each truck and how many trucks should have a fleet for a given city. The proposed idea consists of having a wireless device in the container, in order to permit a particular communication between them and to gather a series of data that can be manipulated and used for allowing a better and intelligent management of the wastes processes. The main aim of the proposed idea is the drastic reduction of CO₂ emissions, due to the shortening of the average path length of waste trucks in urban/suburban area with a strong environmental impact. Involving theoretical and empirical analysis it is possible to design a “minimum path” for each truck, having the possibility to visit each single garbage collector once for a single path, minimizing the fuel consumption of vehicles. The waste containers will be equipped by telecommunication nodes, able to accomplish two

main tasks: 1) Measuring the amount of garbage present in the container, in order to decide whether to signal the presence of material to be collected; 2) Communicate with other nodes, in order to signal the presence of the material and inserting itself as a node in the network of full containers exploiting the concept of Machine-to-Machine (M2M) communications. In this way, if a periodic travel scheduling is decided, the exact number and positions of full containers is known and an optimized travel plan can be decided. In order to minimize CO₂ pollutions, the architecture will be structured as follows: 1) A distributed set of signaling nodes, that will be on or off, depending on the level of existing garbage to be picked up; 2) A set of coordinator nodes, able to collect signaling messages and to communicate to Head Quarter (HQ) the exact topology of the active network (in terms of container positions, waste priority and link cost among two active nodes); 3) A communication protocol, able to manage the coexistence of a plurality of containers and optimize nodes energy consumptions; 4) A set of minimization algorithms, able to take into account the Hamilton and minimum path theories, that give the opportunity of optimizing path length and cost with a consequent CO₂ emission reduction thanks to reduction of unnecessary dumpster truck cruising, leading to fewer vehicles on the road, as well as less air and noise pollution. The solution of providing the waste collection system with newest technologies in order to permit real-time bidirectional communication between the on-road infrastructures and the service operators/managers. Moreover, it Potentially helps to avoid exceptional incidents such as a fire in the containers by facilitating real-time response. The main research topics will be focused on different telecommunication issues. After the structure of the network that has to be taken into account is modeled, the appropriate communication technology should be investigated. Taking into account the overall distance to be covered, the mean signaling period and the allowed transmission levels, a mesh topology is strongly desired if a full coverage of the geographical region needs to be obtained. So it is important to investigate what is the best standard to be applied. In addition, users behaviors and garbage levels need to be observed, with the aim of discovering whether a periodical signaling is suitable or not, on the basis of the frequency of truck trips. Nodes energy consumption need to be considered, as well as the number of times and the moments the container nodes need to be recharged (we assume that

waste trucks can restore the correct energy level of nodes when needed). Finally, the communication protocol, integrated with optimization algorithms, should be designed, taking into account the opportune metrics, based on energy optimization and travel distance minimization (with the consequent CO₂ reduction). Regarding the practical development of the proposal idea, once the technology has been investigated and chosen, the signaling nodes will be installed on the sites and the final integration software will be able to manage the overall operations, giving the opportunity to HQ (where the trucks are queued for service) to a priori know what is the “best” route to be followed. With the deployment of a fill status monitoring solution in the urban and sub-urban areas, there is the opportunity to develop an optimization framework for the waste collection routes. The first challenge is to devise and implement an architecture to store and retrieve, when necessary, the information obtained from the container fill sensors. This includes specifying the information workflow, the database schema and formats to exchange data between modules. The second goal is to analyze and compare different algorithms for the optimization of waste collection routes so that an efficient itinerary can be calculated within a timeframe of two hours, given the containers fill status and the collection vehicles capacities

REFERENCES

- [1] A. Moustafa, A.A. Abdelhalim, A.B. Eltawil, N. Fors (2013). Waste Collection Vehicle Routing Problem: Case Study In Alexandria, Egypt *The 19th International Conference on Industrial Engineering and Engineering Management*.
- [2] Christos Chalkias, Katia Lasaridi (2009). A GIS based model for the optimisation of municipal solid waste collection: the case study of Nikea, Athens, Greece *WSEAS Transactions on Environment and Development*, Issue 10, Volume 5, October.
- [3] Optimizing Efficiency, Economy, and Traceability in Waste Management *Technology Basics, White Paper*, 2012 HID Global.
- [4] N.P. Thanh, Y. Matsui, N.V.C. Ngan, N.H. Trung, T.Q. Vinh and N.T.H. Yen (2009). GIS application for estimating the current status and improvement on municipal solid waste collection and transport system: Case study at Can Tho city, *Vietnam Asian Journal on Energy and Environment*.
- [5] Danijel Markovi, Dragoslav Janoevi, Miomir Jovanovi, Vesna Nikoli (2010). Application Method for Optimization in Solid Waste Management System in the City Of NI *Mechanical Engineering* Vol. 8, No 1, 2010, pp. 63–76.

A new application for analyzing driving behavior and environment characterization in transportation systems based on a fuzzy logic approach

P. Fazio, A.F. Santamaria, M. Tropea, A. Serianni & F. Cirillo
D.I.M.E.S. Department, University of Calabria, Rende, Italy

ABSTRACT: The physical security in transportation systems is becoming a serious issue in the last years, given the high number of accidents and emergency situations. With the huge arising of technological applications availability in vehicular environments, many efforts have been given in the scientific world, aimed at minimizing the probability of road accidents. In this paper, we propose a new platform able to discover dangerous driving behaviors. We based our application on the on-board diagnosis standard, able to provide all the needed information directly from the electronic control unit of the vehicle. We integrated the received data with a fuzzy logic approach, obtaining a description of how the driver is behaving. The overall system can take several initiatives (alarms, rpm corrections, etc.), in order to notify the bad behavior of the driver. The performance of the proposed scheme have been validated through a deep campaign of driving simulations.

Keywords: vehicular networks, fuzzy logic, driving behavior, smart device, OBD

1 INTRODUCTION

During the last decade, the interest for vehicular opportunities and potentialities has heavily grown, due to the numerous advantages given by the technological progress. In particular, road safety has been one of the major objectives to be reached. Road accidents represents the main cause of deaths each year: in most cases they are due to reckless driving styles employed by drivers (pedal pressures, steering speed, etc.). The real-time identification of potentially dangerous driving styles is an important element for road safety, as it gives the opportunity to take precautions in terms of safety distance, speed, etc. The main goal of this work is represented by the characterization of the different driving styles in various environments, with the possibility of highlighting potentially dangerous behaviors. Our approach is based on the deployment of a Smart-Device (SD, a phone, a tablet, etc.) to acquire, process data and perform the characterization (nowadays, at least one SD is present in a vehicle). The SD gives the possibility to acquire information from the Internet (e.g. weather data), from its own sensors (such as the gyroscope), from the built-in GPS and from the Electronic Control Unit (ECU) via On Board Diagnostics II (OBD-II) standard [1], [2], [3]. In particular, the SD allows to interface with the vehicle OBD-II via Bluetooth. There are many dedicated devices, like the KiWi

BlueTooth (KW-BT) [4], able to read data from the ECU of the most of the vehicles. Generally, inside the KW-BT device, there is an ELM327 microcontroller, able to take data from the OBD-II port and encapsulate it for the transmission via Bluetooth. In this paper, the use of the fuzzy logic paradigm [5], [6] allows us to analyze different continuous variables, dynamically evaluating the their degree of membership (also called as degree of truth) to different fuzzy sets. Unlike the binary logic, the fuzzy one have no “static” boundaries for its sets, but include a variation of a threshold value that is as an approximation of a subjective judgment. That is to say, the degree of membership of an object to a fuzzy set can take any value between [0,1], unlike the traditional concept, which is restricted only to the values 0 and 1. In particular, for our purposes and to identify a particular driving style, we first need to pick out the environment in which the driver is located (such as urban, suburban or highway). The identification can be made by performing a statistical analysis on the speed data, acquired experimentally from the SD and the OBD-II. For this purpose, after acquiring a large number of samples for different environments, a Gaussian statistical analysis has been introduced, in order to have information about average speeds and variance, recorded in different intervals. After a particular environment is identified, the proper rules to decide the driving styles are associated to it. In our

work, the term “driving style” refers to the way a driver lead the vehicle. In our work, we started from some existing studies [7], [8], about driving styles recognition and, then, we refined them by taking into account the characteristics of different road environments (urban, suburban or highway). It is important to observe that, in different environments with equal speed and acceleration, the same behavior can be considered in different ways. For example, a driver who travels at 35 km/h in urban environment is classified with normal behavior, but if the speed of 35 km/h is maintained on the highway, the behavior is no longer normal, but it can indicate the presence of traffic jams or abnormal situations. For this reason, it is not possible to use the same fuzzy speed sets in all environments (an analysis will be made for each case). After the implementation of the Android application, all the analyses have been carried out through the Fuzzy Logic and Statistical Toolboxes of MATLAB. This paper is organized as follows: Section II presents an in-depth overview on state-of-the-art of similar approaches in VANETs; Section III introduces the considered scenario, while Section IV offers a deep description of the proposed scheme. Section V validates the proposed scheme and, then, conclusions are summarized in the last section.

2 RELATED WORK

There are many works in literature based on some criteria or schemes for recognizing how a driver is behaving during a trip on a road. In [9] and [10] the authors underline the importance of controlling a vehicle and how it reflects on fuel economy and emission reduction, since they are heavily influenced by road conditions and driving styles. Their main idea consists in predicting future road and environmental conditions in order to a-priori know how a driver should act to optimize consumptions. The authors state also that each individual driving style is different and it is not so easy to meet the optimal driving conditions. So, another important effort could be given by investigating the driving style factors that have a major impact on fuel economy. In [11] the importance of identifying driving styles is underlined, especially for Intelligent Transportation Systems (ITS). A driving style identification is proposed, based on the concept of normalizing driving behavior using a personalized driver modeling. The authors use a neural network to learn driver features, then the obtained model is tested and used for defining an aggressiveness index, able to discover abnormal driving behavior. Another important contribution is given by the work in [12], in which the effectiveness of an accurate driving style recognition is underlined. Advanced driving assistant systems (adaptive

cruise control systems, intelligent forward collision warning systems, platooning warnings, etc.) represent the future in ITS, so the understanding of driving behaviors is very crucial. In their work, the authors exploit the potentialities of a clustering method for the elaboration and analysis of the commonness and individuality of driving behavior characteristics, extracting data from the ECU of the vehicle. In addition, since the gathered data is very large, a data mining approach has been also employed for the extraction of deeper information. The work proposed by Vaitkus et al. in [13] is based on the inertial measurement signals of the vehicle with the help of GPS. The proposed monitoring system is capable to classify aggressive and normal driving styles, by applying a pattern recognition approach, with a relatively high success rate. In the work presented in [14], driver behaviors are rated in common traffic conditions and, then, through a statistical analysis of the collected data, the driver is characterized as aggressive, anxious, keen, or other. In this way, assistance services can be realized to be personalized to the particular driver. The authors of [15] proposed a scheme based on a smart-phone application for detecting and classifying driving maneuvers, using the smart-phone accelerometer and vehicle gyroscopes. The work proposed in [16] used different classifying techniques to detect driving events from smart-phones using inertial sensors as well as directly from the ECU of the vehicle. They also made a deep comparison of the obtained data from both sources. Also in [17], the authors propose a scheme for identifying a driver by inertial sensors. Their study is based on the possibility of separate data into classes, considering the basic events of accelerating, braking, and turning. All the possible maneuvers can be considered to be a composition of the basic events. Differently from the cited works, in our proposal we considered the fuzzy approach and the modern OBD-II connection, from which an application on a generic SD can gather the appropriate data. In particular, the main contributions are: a) Study and development of a fuzzy approach to be applied to vehicular environments, in order to obtain a scheme able to recognize the way the driver is behaving during a trip in a given environment; b) Statistical analysis of users average speed in different topologies and environments; c) Definition of fuzzy subsets on the basis of the pdf statistics of the particular road.

3 THE ON-BOARD DIAGNOSTICS, THE CONTROLLER AREA NETWORK (CAN) PROTOCOL AND THE FUZZY LOGIC PARADIGM

In this paragraph, some details about the considered components of our idea are given.

3.1 The On-Board Diagnostics II

The term On-Board Diagnostics [1], [2] is referred to the ability of the vehicle of self-diagnosing and error reporting. OBD systems give access to the information on the “health state” of the various “emission-relevant” subsystems of the vehicle, such as catalyst, oxygen sensors, while other systems (e.g. air bags, air conditioning, etc.) does not have a self-standard, so any car manufacturer can adopt its own decisions. On-Board Diagnostics born in the late 60 s and early 70 s in America, when the problem of air pollution due to vehicle emissions arisen. In those years, car manufacturers were beginning to install the on-board electronic equipment to check the status of the vehicle. In 1970, a congress approved the Clean Air Act and established the Environmental Protection Agency (EPA), which established some standard levels of maximum permitted emissions and the related maintenance to be done to vehicles to reduce emissions. The OBD is able to acquire by a wired interface the read-only diagnostic signals and data in real-time from all the vehicle control units.

In Europe this standard was introduced for gasoline engines in 2001 along with the emission level Euro 3 with Directive 98/69/EC and it is usually called E-OBD (European OBD) [8]. The standard OBD-II [1], [2] specifies the type of diagnostic connector and its pin, the available electrical signaling protocols, the message format, and a list of monitoring parameters. The standard also provides an extensible list of DTCs (Data Trouble Codes), i.e. the error codes in the standard format, which can be interpreted by any user. Fig. 1 shows the structure of the OBD-II port with the related pins. As shown, the connector is composed by 16 pins arranged in two rows, and five different communication protocols have been defined, although the majority of vehicles are using only one: Variable Pulse Width (VPW, J1850, proposed by General Motors), Pulse-Width Modulation (PWM, J1850,

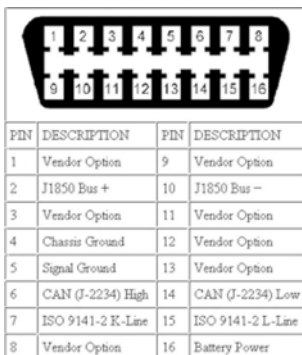


Figure 1. OBD-II connector and pin-out.

proposed by Ford), ISO 9141 (proposed by Chrysler in Asia and Europe), KeyWord Protocol 2000 (KWP2000, ISO 14230) and ISO15765 (via CAN). The OBD-II is no longer used only to diagnose vehicle problems: the provided information could be used by telecommunication systems or SD applications installed on the vehicles to communicate to third parties. All the gathered data can be useful for any kind of application aimed at enhancing the safety and comfort level of the driver.

3.2 The CAN protocol

The Controller Area Network (CAN or CANbus) is a standard for serial bus, introduced in the eighties with the main aim to connect in real-time various Electronic Control Units (ECU). At the moment, it is the most widespread mean of communication in vehicles. CAN was specifically designed to operate also in the environments disturbed by the presence of electromagnetic waves and it uses a dedicated line differentially balanced in potential, such as RS-485. Thanks to its simplicity and robustness to noise, the CAN is now widely used also in the industry.

The ISO 15765-2 [18] is an international standard for sending data packets on the CANbus. The most common application of this protocol is the transfer of diagnostic messages to devices that use the OBD-II. The wide diffusion of the CAN protocol has determined wide availability of chip transceivers, microcontrollers that integrate CAN ports, development tools, as well as a considerable

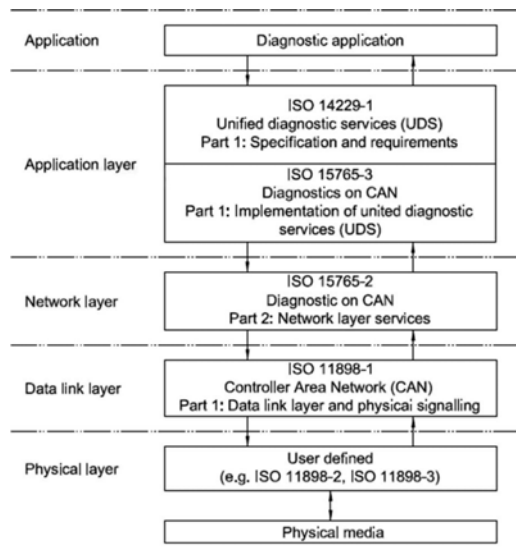


Figure 2. ISO-OSI reference model for CAN bus.

decrease in the cost of these systems. The CAN protocol has an amazing ability to recognize errors and the probability that a message is corrupted and/or not recognized is practically null. All the protocols defined for the CAN bus stack are illustrated in Fig. 2. The communication, in the CAN bus, takes place via different devices, such as sensors or actuators, capable of producing data independently. In addition, this type of equipment, is able to request and use the data produced by any another device. The CAN bus provides the “multi-master” feature, i.e. all nodes of the network can transmit and request the transmission channel simultaneously. For more details about the CAN protocol, please refer to [18].

3.3 The Fuzzy Logic (FL) paradigm

The FL challenges and changes the concept of binary logic (only two states): in the real world everything is a matter of measure, not only white or black, but also shades [19]. Unlike the binary logic, to allow a greater relationship with the natural language, the fuzzy sets do not provide “hard” boundaries but include a landmark change in the considered values. In this way a good approximation to the subjective judgment can be reached. This is why in FL some linguistic variables are used (such as “very”, “somewhat”, “a little”, etc.) to facilitate the expression of rules and facts. The linguistic variables are coded with appropriate functions. This concept is summarized in Fig. 3.

The Membership Degree (MD) of an object referred to a fuzzy set can assume any value in the range $[0,1]$, unlike a traditional set, which is restricted to the values 0 and 1 (false and true): in FL, the MD is to be intended as indicating “how much” a property is true. FL systems are based on the IF-THEN (antecedent-consequent) rules, without the ELSE part. Through some input-output relationships it is possible to approximate any function or system to describe or control. One of the most usual inference method is the Mamdani approach [20], divided into four main steps: input fuzzyfication, inference rule evaluation, aggregation and defuzzyfication. The other one is the Sugeno method [21]; the author suggested the use of a single value (singleton) as a membership function. A singleton is a fuzzy set with a mem-

bership function that is unitary at a particular point and zero otherwise. The Mamdani method is generally used to describe the knowledge and the experience in an intuitive way, while the Sugeno approach is efficient and it is used in optimization problems or adaptive control. Let X be the universe of the considered event, and let x be its elements. At the base of FL there is the Linguistic Variables (LVs) theory: a LV can assume its values as linguistic terms. For example, if we consider the “speed” variable, its universe X could be the set $[0, 300]$ km/h, while the FL linguistic sub-sets could be ‘very slow’, ‘slow’, ‘medium’, ‘fast’, ‘very fast’. Following the classical theory, the set A is defined on X by the Characteristic Function (CF) $f_A(x)$ of the A set:

$$f_A(x) = X \rightarrow \{0,1\}, \text{ where } f_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} \quad (1)$$

and, clearly, it represents a map of the universe X to the set $\{0,1\}$. In the FL paradigm a Membership Function (MF) $\mu_A(x)$ is used instead:

$$\mu_A(x) = X \rightarrow [0,1],$$

$$\text{where } \mu_A(x) = \begin{cases} 1 & \text{if } x \text{ completely in } A \\ 0 & \text{if } x \text{ not in } A \end{cases} \quad (2)$$

$$0 < \mu_A(x) < 1 \text{ if } x \text{ is partially in } A$$

and, clearly, it indicates the “membership amount” of x to A .

4 ENVIRONMENT IDENTIFICATION AND DRIVING BEHAVIORS

In order to characterize the different driving styles it is necessary to in-advance identify the environment in which the driver is located. A human driver instantaneously recognizes the context that surrounds it but, in order to automate the detection of a possible aggressive driving behavior, it is necessary to identify the characteristics of the current environment. In this work, we decided that characteristic parameter for each environment is the average speed, after many empirical observations, that have been made on the obtained experimental data. It is well known that the average speed maintained in a urban environment is very different from the one maintained in the extra-urban one and/or motorway, due to the intrinsic topological nature of the considered roads, as well as the objective constraints that should be respected (speed limits, and traffic lights or pedestrians if present, etc.). By the deployment of different Android APP components that have been developed by us, it is pos-

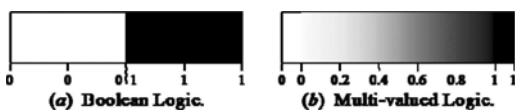


Figure 3. Main difference between Boolean logic and FL.

sible to acquire vehicle dynamics, directly via the KW-BT interface (fuel consumption, acceleration/ deceleration, torque, etc.). Fig. 4 illustrates some demo screens of the considered apps.

The APP is able to send the collected information to a remote server (via HSPA or WLAN if available), in order to perform an off-line elaboration. To this aim, all data were collected using the “01-show current date” mode, as defined in SAE J1979 standard [22], and using the PIDs shown in Table 1. More details can be found in [22].

We focused our attention on the ‘vehicle speed’ field, in order to discern the various types of results that could be achieved. In particular we provided to deeply analyze the obtained samples for different environments and, excluding the zero values (when then vehicle stops), all the AVerage Speed (AVS) values have been observed to follow a Gaussian distribution. So the general expression of the AVS pdf can be considered to be:

$$f_{AVS}(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}}, \quad (3)$$

where μ and σ are, respectively, the average and standard deviation.

Table 1. An extract of all PIDs in SAE J1979.

PID (hex)	Data bytesreturn	Description	Minvalue	Maxvalue	Units
0C	2	Engine RPM	0	16,383.75	rpm
0D	1	Vehicle speed	0	255	km/h
0F	1	Intake air temperature	-40	215	°C
10	2	MAF air flow rate	0	655.35	grams/sec
11	1	Throttle position	0	100	%
1F	2	Run time sinceengine	0	65,535	seconds
2F	1	Fuel Level Input	0	100	%



Figure 4. Some screenshots of the developed APP for car life monitoring.

It is possible to evaluate the error of the considered average AVS, based on confidence intervals/ levels, considering the worst case error probability ξ . It is possible to select a T_{AVS} for a mobile host so that:

$$Prob(CPT < T_{AVS}) < 1 - \xi. \quad (4)$$

The T_{AVS} is called a $(1-\xi)*100\%$ upper confidence bound for average AVS. The assumption of a Gaussian pdf has been verified through the Kolmogorov-Smirnov (KS) normality test [23]. MATLAB gives the opportunity to simply analyze the considered data, using the *kstest* function. The cumulative distribution function (cdf) of the average AVS from eq. 1 is:

$$F(t) = P(X_{AVS} \leq t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \quad (5)$$

and the probability that AVS is lower than a value t with a fixed error threshold ξ is:

$$P(AVS \leq t) = P\left(Z \leq \frac{t - \mu_{AVS}}{\sigma_{AVS}}\right) = \Phi\left(\frac{t - \mu_{AVS}}{\sigma_{AVS}}\right) = 1 - \xi \quad (6)$$

where Z is a random variable equal to $(AVS - \mu_{AVS}) / \sigma_{AVS}$. The $\Phi(\cdot)$ function represents the standard Gaussian distribution function. At this point, through the tabular values of the standard normal distribution, it is possible to obtain the AVS estimation for a given threshold, such as referred in [24], [25]. Clearly, the values of μ and σ are strictly dependent on the topology and mobility conditions. Road accidents are the cause of numerous deaths each year and, in most cases, the blame is attributed to dangerous driving attitudes taken by drivers. Preventively identifying a driver who is adopting a way of potentially dangerous driving is an important element in road safety, as it gives the opportunity to take precautions. The risk of accidents increases with the speed of the vehicle, but additional risk factors are represented by risky maneuvers, such as frequent and/or sudden lane changes, or driving into altered states. With the term “driving style” we refer to the way a person prefers (or is used) to guide [26]. In the literature there are several studies that will identify the different driver attitudes by analyzing different data drivers. Some of them use the recognition of facial features to detect driver fatigue, other ones analyze the movements of the steering wheel [27], others identify a drunk driver according to data provided by the accelerometer of a smartphone which are compared with experimental data [28]. In [29] the authors used smartphone sensors to recognize

the different driving styles. Our analysis is based on some studies ([30], [31], [32]) in which four riding modes (Below Normal, Normal, Aggressive, and Very Aggressive) are derived, using as input data the Euclidean norm, the transverse acceleration and vehicles speed. The use of the Euclidean norm of the acceleration in two dimensions allows to obtain more accurate information regarding the driving behavior taken along the way. The norm is calculated by eq. 7 for each point, then the signal is averaged by using a window of N samples, as in eq. 8:

$$Norm(n) = \sqrt{Accel_{Long}^2(n) + Accel_{Transv}^2(n)} \quad (7)$$

$$Norm(i) = \frac{1}{N} \sum_{n=1}^N Norm(n) \quad (8)$$

The acceleration rule has been divided into three fuzzy sets (Low, Medium and High), while the speed rule in five sets (Very Low, Low, Medium, High and Very High). Unlike the effort in [30], in our work we used trapezoidal functions. We made a characterization of driving styles in different road environments, according to the current law permissions in Italy. As regards the data relating to the standard acceleration, as illustrated in Fig. 5, the membership functions described in [30] have been maintained, as well as the 15 fuzzy rules to identify the driving style.

From Fig. 6 we can observe that the x axis indicates the target behavior, with a value between 0 and 1: the scale indicates the level of aggressiveness of the driver. A value between 0 and 0.2 indicates a Below Normal (BN) behavior (generally too low speed for a given environment), a value of 0.4 indicates a Normal behavior, a value of

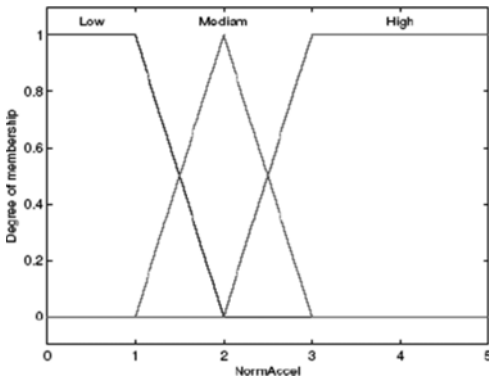


Figure 5. The membership functions considered for the acceleration.

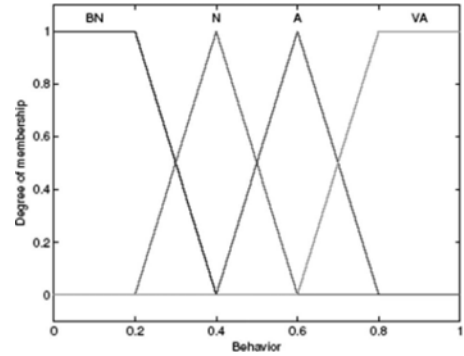


Figure 6. The membership functions considered for the driving style recognition.

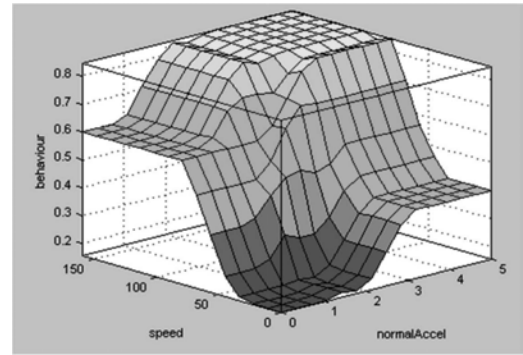


Figure 7. Driving behavior trend in function of the considered inputs (speed and normal acceleration).

0.6 indicates an Aggressive behavior, and values equal or higher 0.8 indicate a very aggressive behavior. The intermediate values indicate the transition behaviors (for example, a value of 0.55 indicates a behavior that is a bit more than normal but a little less than aggressive and so on). Given that the considered system is a function of two variables, its output is represented with a three-dimensional graph, as shown in Fig. 7, in which the different chromatic gradations highlight the transaction, in terms of membership value, from a fuzzy set to another.

It must be underlined that the collection of values from sensors is everyday applied in different environments [33–35], in order to enhance the offered services. The use of GPS and radar, especially in vehicular environment [36], [37], are used widespread, for positioning and security applications.

5 PERFORMANCE EVALUATION

In this subsection, the main measurements and results are shown. First of all, different speed samples have been collected through the analysis of the OBD-II data, obtained by different trips in different considered maps. Then, the data has been analyzed through the previously proposed environment recognition scheme. In particular Fig. 8 shows the considered paths (urban on the left, extra-urban on the right). Next tables illustrate the trend of the Gaussian parameters for the average speed in a urban scenario. It is observed that for the same speed, the recognized behavior is considered differently depending on the environment (remember that the values from 0 to 0.2 indicate a “below-normal” behavior, a value of 0.4 indicates a “normal behavior”, a value of 0.6 is an “aggressive behavior” and values equal or higher than 0.8 indicate a “very aggressive” behavior). The mean values and pdf trends are summarized in Fig. 9.

Figure (Fig. 10) illustrates the main window of the MATLAB application used for evaluating the correctness of our scheme.

The last table illustrates the main obtained results. All the results are obtained by deploying the membership function illustrated in Fig. 11.

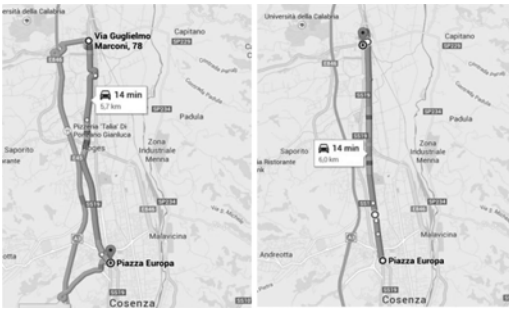


Figure 8. Two different scenario for measuring the speed trend of vehicles.

Table 2. Speed for Urban Scenario (early morning).

SAMPLE	MEAN [km/h]	VARIANCE [(km/h) ²]	STANDARD DEVIATION [km/h]
M1	27.9351	443.086	21.04960807
M2	27.34	344.179	18.55206188
M3	27.9141	516.256	22.72126757
M4	24.6083	207.978	14.42144237
M5	27.1502	277.699	16.66430317
M6	25.4977	292.616	17.10602233
M7	29.6707	587.55	24.23943069
M8	27.5908	231.299	15.20851735
M9	24.2135	232.393	15.24444161
M10	27.9529	158.837	12.60305519

Table 3. Speed for Urban Scenario (lunch time).

SAMPLE	MEAN [km/h]	VARIANCE [(km/h) ²]	STANDARD DEVIATION [km/h]
PR1	34.216	474.393	21.78056473
PR2	30.2456	220.068	14.83468908
PR3	33.1402	478.109	21.86570374
PR4	33.5199	319.969	17.88767732
PR5	30.5065	155.852	12.48406985
PR6	31.7677	638.157	25.26176953
PR7	32.2651	337.468	18.37030212
PR8	31.0514	193.484	13.90985262

Table 4. Speed for Urban Scenario (late afternoon).

SAMPLE	MEAN [km/h]	VARIANCE [(km/h) ²]	STANDARD DEVIATION [km/h]
PM1	32.7217	419.048	20.47066193
PM2	23.3464	389.436	19.73413287
PM3	26.7461	395.122	19.87767592
PM4	31.5702	134.79	11.60990956
PM5	31.9628	174.05	13.19280107
PM6	34.9483	155.916	12.48663285
PM7	24.0489	359.53	18.96127633
PM8	30.1541	484.582	22.0132233
PM9	29.6787	149.951	12.24544813
PM10	28.9043	383.131	19.5737324
PM11	37.0024	321.583	17.93273543
PM12	30.9663	208.509	14.43984072

Table 5. Speed for Urban Scenario (evening).

SAMPLE	MEAN [km/h]	VARIANCE [(km/h) ²]	STANDARD DEVIATION [km/h]
S1	32.165	73.4429	8.569883313
S2	37.0315	231.765	15.22383
S3	34.4443	186.414	13.65335124
S4	32.8217	124.87	11.1745246
S5	32.2918	72.9507	8.541118194
S6	31.8611	109.42	10.46040152
S7	28.7028	255.974	15.99918748
S8	34.6001	154.901	12.44592303
S9	28.7734	203.79	14.27550349

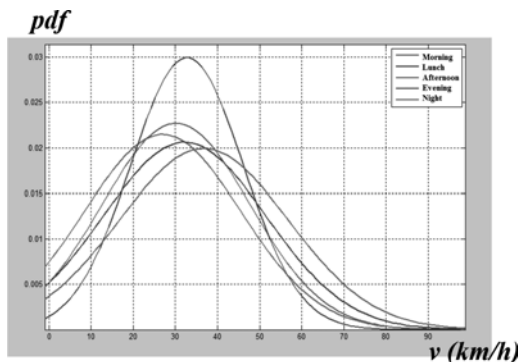


Figure 9. Gaussian approximations for the average driving speed in different periods of a day.

The terms of Table 6 are Urban (U), Extra-Urban Secondary (EXS), Extra-Urban Primary (EUP) and Highway (A).

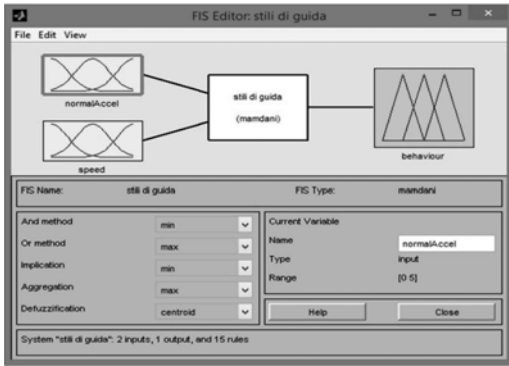


Figure 10. Main MATLAB tool window used for evaluating the correctness of the proposed idea.

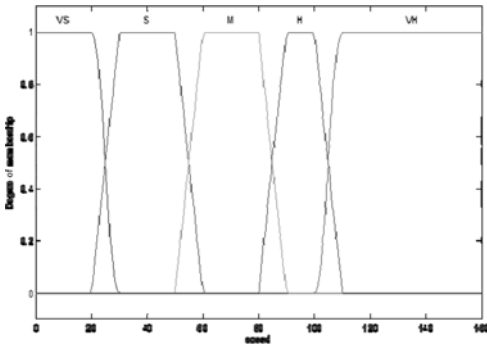


Figure 11. Speed membership function used for an extra-urban environment.

Table 6. The Main Obtained Results.

Acceleration norm [m/s ²]	Speed [km/h]	Recognized behavior			
		U	EXS	EXP	A
0.5	38	0.4	0.153	0.153	0.153
0.5	70	0.6	0.4	0.4	0.153
0.6	90	0.6	0.6	0.4	0.4
0.6	110	0.6	0.6	0.6	0.4
2	75	0.847	0.6	0.6	0.153
4	55	0.847	0.5	0.4	0.4
3	75	0.847	0.6	0.6	0.4
3	60	0.847	0.6	0.4	0.4

6 CONCLUSIONS

In this paper a new algorithm for characterizing driving behaviors is proposed. It is based on the CAN protocol and OBD-II standard, while the decision is taken by the use of a FL approach. The choice of FL (in place of the traditional binary logic) has been dictated by the needing to manage continuous ingress/egress variables, evaluating the effects on the

output for a given input. We identified four riding modes and to each of them a level of aggressiveness has been assigned. The intermediate values represent transient conditions from one style to another one. The classification of driving styles was made for each different environment, identified by a set of Gaussian curves, related to the average speed of the vehicle. The data were collected using an application developed for Android OS, which is able to acquire the data of the vehicle speed and acceleration. Many numerical results have been obtained, as shown in the tables and the average accuracy in revealing mobile hosts behavior is around 85%. So, in the most of the cases, the application is able to correctly identify the driving behavior of the user.

REFERENCES

- [1] Wonang Jang, Daeseong Jong and Dohoon Lee, "Methodology to improve driving habits by optimizing the in-vehicle data extracted from OBDII using genetic algorithm," *2016 International Conference on Big Data and Smart Computing (BigComp)*, Hong Kong, China, 2016, pp. 313–316.
- [2] K. Smith and J. Miller, "OBDII data logger design for large-scale deployments," *Intelligent Transportation Systems - (ITSC), 2013 16th International IEEE Conference on*, The Hague, 2013, pp. 670–674.
- [3] Geraldo, G., "Differences between On Board Diagnostic Systems (EOBD, OBD-II, OBD-BR1 and OBD-BR2)," SAE Technical Paper 2006-01-2671, 2006.
- [4] http://www.jscspeed.com/catalog/Kiwi_Bluetooth_OBDII_for_Android-30126-1.html
- [5] N. Abbas and J. J. Saade, "A fuzzy logic based approach for network selection in WLAN/3G heterogeneous network," *Consumer Communications and Networking Conference (CCNC), 2015 12th Annual IEEE*, Las Vegas, NV, 2015, pp. 631–636.
- [6] D. Ramot, M. Friedman, G. Langholz and A. Kandel, "Complex fuzzy logic," in *IEEE Transactions on Fuzzy Systems*, vol. 11, no. 4, pp. 450–461, Aug. 2003.
- [7] R. Wang and S. M. Lukic, "Review of driving conditions prediction and driving style recognition based control algorithms for hybrid electric vehicles," *Vehicle Power and Propulsion Conference (VPPC), 2011 IEEE*, Chicago, IL, 2011, pp. 1–7.
- [8] D. Dörr, D. Grabengieser and F. Gauterin, "Online driving style recognition using fuzzy logic," *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*, Qingdao, 2014.
- [9] Rui Wang, Lukic, S.M., "Review of driving conditions prediction and driving style recognition based control algorithms for hybrid electric vehicles", *Vehicle Power and Propulsion Conference (VPPC), 2011 IEEE*, pp. 1–7, DOI: 10.1109/VPPC.2011.6043061.
- [10] Malikipoulos, A.A, Aguilar, J.P., "Optimization of driving styles for fuel economy improvement", *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on*, pp. 194–199, DOI: 10.1109/ITSC.2012.6338607.

- [11] Shi, B., Xu, L., Hu, J., Tang, Y., Jiang, H., Meng, W., Liu, H., "Evaluating Driving Styles by Normalizing Driving Behavior Based on Personalized Driver Modeling", *Systems, Man, and Cybernetics: Systems*, IEEE Transactions on, Year: 2015, Volume: PP, Issue: 9, DOI: 10.1109/TSMC.2015.2417837.
- [12] Geqi Qi, Yiman Du, Jianping Wu, Ming Xu, "Leveraging longitudinal driving behaviour data with data mining techniques for driving style analysis", *Intelligent Transport Systems, IET*, 2015, Volume: 9, Issue: 8, pp. 792–801, DOI: 10.1049/iet-its.2014.0139.
- [13] Vaitkus, V.; Lengvenis, P.; Zylius, G., "Driving style classification using long-term accelerometer information", *Methods and Models in Automation and Robotics (MMAR)*, 2014 19th International Conference On, pp. 641–644, DOI: 10.1109/MMAR.2014.6957429.
- [14] Bar, T., Nienhuser, D., Kohlhaas, R., Zollner, J.M.; "Probabilistic driving style determination by means of a situation based analysis of the vehicle data", *Intelligent Transportation Systems (ITSC)*, 2011 14th International IEEE Conference on, pp. 1698–1703, DOI: 10.1109/ITSC.2011.6082924.
- [15] D.A. Johnson, M.M. Trivedi, "Driving style recognition using a smartphone as a sensor platform", in *Intelligent Transportation Systems (ITSC)*, 2011 14th International IEEE Conference on, pages 1609–1615. IEEE, 2011.
- [16] A. Sathyanarayana, S.O. Sadjadi, J.H. Hansen," Leveraging sensor information from portable devices towards automatic driving maneuver recognition", in *Intelligent Transportation Systems (ITSC)*, 2012, 15th International IEEE Conference on, pp. 660–665.
- [17] W. Shi, J. Yang, Y. Jiang, F. Yang, Y. Xiong, "Senguard: Passive user identification on smartphones using multiple sensors", in *Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2011 IEEE 7th International Conference on, pages 141–148.
- [18] Road vehicles—Diagnostics on Controller Area Networks (CAN)—Part 3: Implementation of unified diagnostic services (UDS on CAN), *International Standard ISO 15765-3*, 2004.
- [19] L. A. Zadeh, "Fuzzy logic: issues, contentions and perspectives", *Acoustics, Speech, and Signal Processing*, 1994. ICASSP-94., 1994 IEEE International Conference on, vol. 6/183.
- [20] Mamdani, E. H. (1977). Application of fuzzy logic to approximate reasoning using linguistic synthesis, *IEEE Transactions on Computers* 26(12): 1182–1191.
- [21] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE transactions on systems, man, and cybernetics*, vol. 15, no. 1, pp. 116–132, 1985.
- [22] http://standards.sae.org/j1979_201202/.
- [23] C. Montgomery, "Applied statistics and probability for engineers", Third Edition, Wiley, 2003.
- [24] J.Banks, J.S. Carson et al., "Discrete-Event system simulation," Third Edition, Prentice Hall, 2001.
- [25] M.A. Stevens, R.B. D'Agostino, "Goodness of Fit Techniques", Marcel Dekker, New York, 1986.
- [26] Laila M. Martinussen, Mette Møller, Carlo G. Prato, "Driver Style And Driver Skill – Clustering Sub-Groups Of Drivers Differing In Their Potential Danger In Traffic", 16th Road Safety on Four Continents Conference, 15–17 May 2013, Beijing, China.
- [27] U.T. Krajewski, D. Sommer and M. Golz, "Steering wheel behavior based estimation of fatigue," in *The 5th international driving symposium on human factors in driver assessment, Training and vehicle design*, June 2009, pp. 118–124.
- [28] J. Dai, J. Teng, X. Bai, Z. Shen, and D. Xuan, "Mobile phone based drunk driving detection," in *Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, 2010 4th International Conference on- NO PERMISSIONS, march 2010, pp. 1–8.
- [29] Minh Van Lyy, Sujitha Martiny and Mohan M. Trivediy, "Driver Classification and Driving Style Recognition using Inertial Sensors", 2013 IEEE Intelligent Vehicles Symposium (IV), June 23–26, 2013, Gold Coast, Australia.
- [30] Ahmad Aljaafreh, Nabeel Alshabat, Munaf S. Najim Al-Din, "Driving Style Recognition Using Fuzzy Logic", 2012 IEEE International Conference on Vehicular Electronics and Safety, July 24–27, 2012, Istanbul, Turkey.
- [31] Maen Saleh, Ahmad Aljaafreh, Nashat Albhour, "Fuzzy-Based Recognition Model for Driving Styles", (*IJECS*) International Journal of Electrical, Electronics and Computer Systems. Vol: 16 Issue: 01, September 2013.
- [32] Ahmad Aljaafreh, "Web Driving Performance Monitoring System", *World Academy of Science, Engineering and Technology* Vol:6 2012–10–28.
- [33] F. De Rango, N. Palmieri, S. Ranieri, "Spatial correlation based low energy aware clustering (LEACH) in a wireless sensor networks," in *Advances in Electrical and Electronic Engineering*, Vol. 13, Issue 4, 2015, pp. 350–358.
- [34] D. Amendola, F. De Rango, K. Massri, A. Vitaletti, "Efficient neighbor discovery in RFID based devices over resource-constrained DTN networks," in *IEEE International Conference on Communications, ICC 2014; Sydney, NSW; Australia; 10–14 June 2014*.
- [35] D. Amendola, N. Cordeschi, M. Shojafar, V. Abate, F. De Rango, "Performance evaluation of a multi-frame persistent neighbor discovery strategy based on Sift-distribution in DTN RFID networks," in *Int. Symposium on Perf. Evaluation of Computer and Telecommunication Systems, SPECTS 2014; Monterey, CA; United States; 6–10 July 2014*.
- [36] M.T. Alrefaie, I. Carreras, F. Cartolano, R. Di Cello, F. De Rango, "Map matching accuracy: Energy efficient location sampling using smartphones," in 6th Int. IEEE Conference on Intelligent Transportation Systems: Intelligent Transportation Systems for All Modes, ITSC 2013; The Hague; Netherlands; 6–9 Oct. 2013.
- [37] A.F. Santamaria, C. Sottile, F. De Rango, M. Voznak, "Road safety alerting system with radar and GPS cooperation in a VANET environment," in *Proc. of SPIE – The International Society for Optical Engineering*, Baltimore, MD; United States; 7–8 May 2014.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Trust-based intrusion detection in mobile ad-hoc networks using a dynamic approach for energy-efficient monitoring

Andrea Lupia

DIMES Department, University of Calabria, Rende (CS), Italy

ABSTRACT: MANETs allow mobile nodes communicating to each other using the wireless medium. The security is a key aspect in these networks, because they are infrastructure-less, so external nodes could disturb the communication. Mobile nodes belonging to the network could be compromised, misbehaving during the data transmission, or they could have a selfish behavior due to energy-saving reasons. The detection of these behaviors need an IDS that takes into account the latest interactions between nodes, so malicious or selfish nodes could be detected reactively over time. Monitoring operations increase the energy consumption, so this issue is addressed by this proposal, reducing the energy required by the monitoring system. The results show an improvement in the saved energy, improving the detection performance too.

Keywords: Ad-hoc Networks, Security, Energy, Intrusion Detection, Trust, Monitoring

1 INTRODUCTION

Mobile nodes have limited amount of energy available for communication and the distinctive security issues of Mobile Ad-hoc Networks (MANETs) need additional measures to ensure reliable communication among nodes. These measures, in the form of Intrusion Detection System (IDS), require energy to perform monitoring operations, affecting the overall energy consumption of the nodes. Reducing the energy consumption by using a dynamic monitoring approach that have no negative effects on the detection effectiveness could improve the lifetime of the nodes, keeping optimal performance under the communication point of view. Our proposal concerns the reduction of the promiscuous mode usage of the wireless interface to monitor sent data, through a relationship among the monitoring activity probability and the trust value of the agent node. This relationship is defined by a Dynamic Monitoring Function (DMF), therefore there is a lower probability of monitoring for nodes with high trust values. The energy consumption improves mainly when the network is composed of fair nodes. The organization of this work is the following: the Section 2 describes other proposals related to the security and the energy consumption subjects in MANETs; the trust modeling framework is shown in Section 4; our proposal about the dynamic monitoring, addressing the security and the energy consumption issues in MANETs, is

described in Section 5; the analysis of the performance achieved by our proposal in terms of energy consumption are discussed in Section 6. Finally the Section 7 shows the conclusions.

2 RELATED WORK

Many proposals about MANETs address energy and security issues separately. Security issues in MANETs depend on their distributed nature. They concern the communication between nodes, since external agents could interfere with the transmission and fair nodes could be compromised. Nodes have limited amount of energy, so security measures need to wisely use it. The work proposed by Tan, Li, & Dong (2015) consists in a trust based routing mechanism for securing the Optimized Link State Routing (OLSR) protocol, using fuzzy Petri net for trust evaluation. The proposal uses four functional modules: trust factor collection module, trust evaluation module, recommendation aggregation module and recommendation propagation module. Its aim is the prevention of malicious or compromised nodes inclusion in the discovered routes. The minimum trust value of an intermediate node is the trustworthiness of the path containing it. Some works (De Rango & Marano 2009, De Rango 2009) proposed a trust-based Secure Ad-hoc On-demand Distance Vector (AODV) using incentive cooperation and an

intrusion detection mechanism to penalize selfish nodes. Hidoussi et al. (2015) proposed a centralized IDS to detect selective forwarding and black hole attacks in cluster-based Wireless Sensor Networks (WSNs), with potential intrusions detection based on control packet analysis. Another proposal is based on a collaborative contact-based approach for selfish nodes detection in MANETs (Hernández-Orallo et al. 2014). A watchdog is defined to perform monitoring operations, which consist in overhearing the transmitted packets in order to detect malicious or selfish nodes. Each node in the network runs a local watchdog to detect selfish nodes and new contacts. The acquired information is managed by the diffusion module. The system evaluates positive and negative detections, in addition to the inability to make a decision when a node does not have enough information. Attack analysis could help in improving security measures for their detection. The vulnerability of SIP server is analyzed and a solution based on an IDS was proposed by Voznak & Safarik (2012). A distributed topology of honeypots (monitoring servers masqueraded as production server) was proposed to gather real data about malicious activity in networks (Safarik et al. 2013). Prediction schemes are used for many purposes in MANETs, they could help in intrusion detection too. Actually, a prediction based data dissemination model for Vehicular Ad-hoc Networks (VANETs) was proposed by Kurmis et al. (2014), while Fazio and Tropea (2012) proposed a markovian prediction scheme for resource reservation in mobile wireless networks. Various authors proposed new approaches changing the path metric to achieve different goals. Fazio et al. (2013) used a multiobjective metric with the main aim of reducing the interference in VANETs. Many proposals need reliable routing operations to offer services like VoIP, live video streaming, etc., as proposed by De Rango, Fazio, Scarcello, & Conte (2014). Mobile nodes in MANETs are subject to constraints on the energy available for routing and security operations, therefore many proposals concern approaches with the aim of analyzing and reducing the energy consumption. Epidemic routing is exploited, optimizing the energy consumption and message delivery probability in Delay Tolerant Networks (DTNs) (De Rango, Amelio, & Fazio 2013). The work proposed by Ravi & Kashwan (2015) proposes a new algorithm that uses energy-saving approaches and a hardware circuit to optimize the energy consumption in MANETs. The Span coordination algorithm defines an energy-efficient scheduling technique, the hardware circuit is incorporated to wake up sleeping nodes. Basurra, Vos, Padget, Ji, Lewis, & Armour (2015) propose an energy-efficient

zone based routing protocol through an intelligent rebroadcasting to reduce the overhead. Their proposal uses distributed and parallel broadcasting techniques in order to reduce redundancy and accelerate the path discovery process. An energy-aware routing protocol that takes into account also the link stability was proposed for distributed wireless networks (De Rango, Guerriero, & Fazio 2012). Previous works (Lupia & De Rango 2014, Lupia & De Rango 2015) analyze the route reliability and the total energy consumption due to transmission and monitoring activities. Their are based on the Secure AODV protocol, which offers protection against attacks concerning impersonation and route disruption encrypting protocol packets. Exploiting a Trust Management Scheme (TMS), a reliable communication is achieved. The performance is analyzed by introducing malicious nodes in the network and evaluating many network parameters concerning the packet delivery and the energy consumption. As result, the communication between the nodes remains reliable when the network is under attack. The TMS improves the network reliability, at the cost of an increased energy consumption due to monitoring operations. This proposal is extended by decreasing the energy consumption due to the IDS through the design and the application of a dynamic monitoring function.

3 PROBLEM FORMULATION

The MANET environment is made of mobile nodes. Monitoring operations regarding data transmission are made by using the promiscuous mode of the wireless interface. It increases the energy consumption of the nodes. Nevertheless, an unprotected network cannot guarantee the correct packet delivery because nodes in MANETs are subject to various kind of attacks. Previous works (Lupia & De Rango 2014, Lupia & De Rango 2015) show that the percentage of energy consumed for monitoring activities in MANETs is around 1/3 of the overall wireless interface consumption. Therefore the need of an approach that requires less energy was pointed out. However, the main task of the IDS is the detection of malicious behaviors, so the energy saving must not affect the intrusion detection effectiveness.

4 TRUST RELATIONSHIPS

The concept of trust can be defined as the certainty whereby an agent will perform such an action from the subject point of view (Sun, Yu, Han, &

Liu 2006). The subject is the node requesting the action execution. The agent trust value is computed by monitoring its behavior during the time, the observations done are evaluated in order to establish the agent trustworthiness. The exploited framework assigns trust values in the range $[-1, +1]$. The agents with high trust values have an higher probability to perform the action, while low trust values point out the subject disbelief about the action execution. The mean value of the range in which the trust value is included indicates an uncertainty condition, in which the subject knows nothing about the agent behavior. A trust relationship could be defined as follows:

$$\{subject: agent, action\} \quad (1)$$

The trust value T and the probability that the agent will perform the action p are tied to the relationship in (1). The framework of trust modeling (Sun, Yu, Han, & Liu 2006) defines a trust value based on the entropy. The trust value T depends on the time of the observation, because the agents behavior changes dynamically. The remembering factor ρ is a positive real number with a value less than or equal to 1. It defines the weight of older action observations. The older observations weight is higher when ρ has a higher value, decreasing when the remembering factor value is lower. The trust value of an agent is computed by the subject through direct interactions. The probability p is computed taking into account the observations done as in the following equation:

$$p\{subject: agent, action\} = \frac{1 + \sum_{i=1}^I \rho^{t_c - t_i k_i}}{2 + \sum_{i=1}^I \rho^{t_c - t_i n_i}} \quad (2)$$

The observations of the action are indexed from 1 to I , ρ is the remembering factor, t_c and t_i represent respectively the current time and the time of the observation (in seconds), n_i is equal to 1 for each observation done, k_i has the value of 1 if the observation is successful (i.e. the action was executed), 0 otherwise. p value is 0.5 when no interactions were observed or they are too old. The trust value computation is based on the entropy function defined below:

$$H(p) = -p \log_2(p) - (1-p) \log_2(1-p) \quad (3)$$

The entropy is referred to the uncertainty in the information theory. The equation (3) is exploited to compute the trust value of an agent. It can assume values among -1 and $+1$, a value of $p = 0.5$ coincides with a trust value of 0 (highest uncertainty about the action execution).

$$T = \begin{cases} 1 - H(p), & \text{for } 0.5 \leq p \leq 1 \\ H(p) - 1, & \text{for } 0 \leq p < 0.5 \end{cases} \quad (4)$$

5 PROPOSAL

Our proposal aims at reducing the amount of monitoring operations without weaken IDS effectiveness by using a dynamic approach, with a monitoring probability that depends on the trust value of an agent.

5.1 Dynamic monitoring

The monitoring aim is the observation and control of the progress or quality of some actions executed by an agent over a period of time. In MANETS, it is oriented to routing operations protection, so misbehaviors of nodes will not disrupt the network communication. An energy-efficient approach has to reduce the monitoring activity when it is not needed, namely when the interactions happen among fair nodes, using the promiscuous mode of the wireless interface only when needed. A relationship between the trustworthiness of the agent and the monitoring probability is defined to address this issue. In order to have a reliable IDS, the network status has to be taken into account. The DMF fulfills this goal following these properties:

1. it represents a probability, so a value in the range $[0, 1]$ is assigned to each trust value;
 2. the monitoring probability is 1 when the agent is unknown or distrusted;
 3. the function must be monotonically decreasing;
- The above properties can be mathematically represented as follows:

$$0 \leq \text{DMF}(T) \leq 1, \quad \text{for } T_{\min} \leq T \leq T_{\max} \quad (5)$$

$$\text{DMF}(T) = 1, \quad \text{for } T \leq T_{th} \quad (6)$$

$$\text{DMF}(T_1) \geq \text{DMF}(T_2), \quad \text{for } T_1 > T_2 \quad (7)$$

The term T_{th} is the trust value that an agent has to overtake to be trusted, T_{\min} and T_{\max} are the minimum and the maximum obtainable trust value. The DMF is parametric, since its shape changes dynamically according to various network characteristics. For this proposal purposes, the well-known Beta distribution is used as DMF.

$$\text{Beta}(x; p, q, a, b) = \frac{(x-a)^{p-1} (b-x)^{q-1}}{\text{B}(p, q) (b-a)^{p+q-1}}, \quad (8)$$

for $a \leq x \leq b; p, q > 0$

This distribution is defined in the range of values $[0, 1]$, having four parameters (p and q determining the shape, a and b as bounds). $B(p, q)$ is the beta function.

$$B(p, q) = \int_0^1 t^{p-1} (1-t)^{q-1} dt \quad (9)$$

Distribution functions are monotonically increasing, so the β -DMF is defined as follows:

$$\beta\text{-DMF}(T) = \begin{cases} 1, & \text{for } T < T_{th} \\ 1 - \text{Beta}(T; p, q, T_{th}, T_{max}), & \text{for } T > T_{th} \end{cases} \quad (10)$$

During the simulations performed in the next sections, we have noticed that the obtained data is mostly dependent on the ratio between the parameters p and q . Therefore q to p ratio (Q2PR) defines the relationship among these parameters:

$$\text{Q2PR} = q/p \quad (11)$$

The above relationship has the following properties:

- as Q2PR tends to 0, the β -DMF keeps a monitoring probability of 100% for higher trust values, having a behavior similar to the standard monitoring;
- for Q2PR = 1 and $T = (T_{max} - T_{th})/2$, the β -DMF has the value of 0.5;
- for Q2PR > 1, the monitoring probability resulting from the DMF suddenly decreases when the trust value exceeds the T_{th} value.

In order to avoid a sudden decrease of the DMF as stated in the Q2PR properties, its value is contained among 0 (excluded) and 1.

6 PERFORMANCE EVALUATION

The performance of the proposed approach is evaluated through various simulations. The scenario is defined by a network composed of a source node, a destination node and many neighbors of the source nodes have a path toward the destination. The source sends the data packets to a neighbor until it is trustworthy. If the chosen node becomes distrusted, a new path is computed excluding that node. The parameters used in this scenario are shown in Table 1.

The used energy model is linear, the fixed cost b represents the cost for accessing the channel, and the incremental cost m depends on the packet size. The values of these parameters were empirically obtained (Feeny 2001).

Table 1. Scenario Parameters.

Parameter	Value
Packets transmitted	1200 packets
Packet size	512 bytes
Rate	{0.5, 1, 2, 4} packets/s
D	{0%, 25%, 50%, 75%, 100%}
p	16
Q2PR	{0.125, 0.25, 0.5, 1}
ρ	{0.666, 0.8, 0.9}
m_{prom}	0.388 mW*/s/byte
b_{prom}	136 mW*s
Runs	100

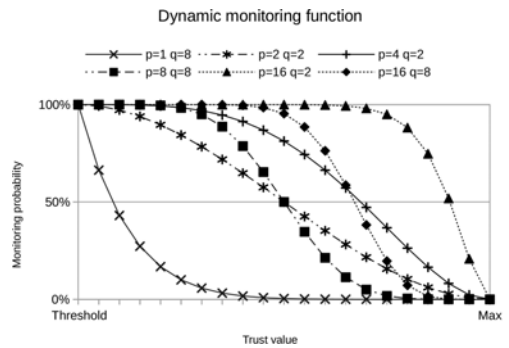


Figure 1. Shape of β -DMF for various p and q values.

$$\text{Cost} = m \times \text{size} + b \quad (12)$$

The 95% confidence interval is shown on the following figures, obtained running many simulations using the same parameter sequences. The channel error model follows a standard uniform distribution, the probability with which the packet transmitted to the intermediate node is lost is the 2%. The parameter D represents the probability that a malicious node will drop a packet to forward.

6.1 Detection accuracy

Standard and dynamic monitoring detect the malicious behavior of a node with $d \geq 50\%$. For lower percentages, the detection is harder, so the DMF parameters need to be tuned consequently. The results of the simulations are shown in Figure 2. Let R represent the transmission rate, in packets per second. The results show that Q2PR have a small effect on the accuracy, but for increasing values it improves a little. With Q2PR = 1, the accuracy of the dynamic monitoring is always equal to or better than the accuracy of the standard monitoring. The parameters R and ρ have a huge effect

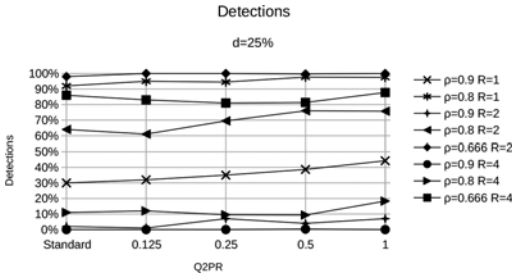


Figure 2. Malicious node detection with $D = 25\%$.

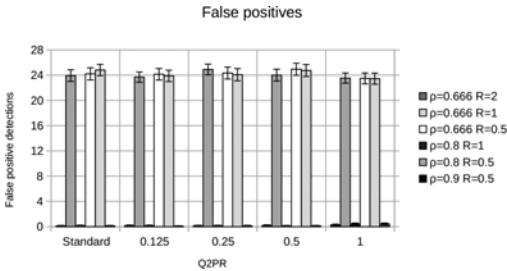


Figure 3. False positive detections.

on the detection accuracy. Lower values of ρ and R allow the detection of small drop percentages. The detection accuracy depends also on the ability to detect only misbehaving nodes without detecting false positive, since the MANET environment brings to lose packets due to the mobility of the nodes and the interferences on the wireless transmission channel. The Figure 3 shows that the false positive detections do not depend on the DMF, keeping almost the same trend for the standard monitoring and the various values of Q2PR.

6.2 Energy consumption

In previous works (Lupia & De Rango 2014, Lupia & De Rango 2015), the energy consumed in monitoring activities was between the 28% and the 41% of the total energy consumption of the wireless interface. This range was limited among the 34% and the 39% for networks composed of fair nodes only. Therefore, saving energy during the monitoring operations improves the nodes lifetime. The energy consumption decreases by using the DMF, as shown in Figure 4. The saved energy is due to the lower amount of monitored packets during fair interactions. Higher values of ρ increases the amount of saved energy, since the average trust value of fair nodes is higher when more positive observations are taken into account, so there is a reduction of the monitoring activity. The energy

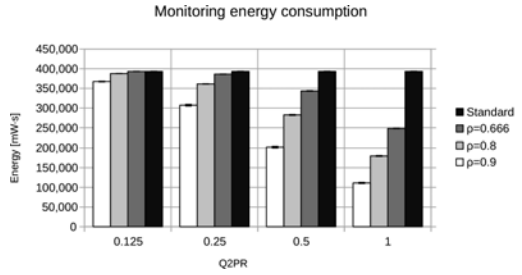


Figure 4. Monitoring energy consumption.

consumption is almost the same of the standard monitoring for lower values of Q2PR, because the amount of monitored packets increases. In the best case, the saved energy is the 70% of the energy consumed while using the standard monitoring, when $\rho = 0.9$ and $Q2PR = 1$.

7 CONCLUSIONS

The dynamic monitoring enables the use of an intrusion detection system also in environments characterized by limited amount of available energy. The proposed approach has also improved the performance of the intrusion detection system, with less time needed to detect misbehaviors, although the main aim of the DMF concerns the energy saving. The detection accuracy is affected by the dynamic monitoring in a positive manner, achieving a little improvement against the standard monitoring. The defined constraints allow maintaining an accurate detection also when the nodes misbehavior is hardly noticeable. Under the energy consumption point of view, the results show a reduction of the consumption of 70% in the best case. The DMF can be tuned according to the network status and the required security level through changing its parameters. The best results were obtained for $Q2PR = 1$, showing just little differences when this ratio is constant and the values of p and q were changed. The definition of the DMF properties allows its implementation through different functions that could be tested in many IDSs, resulting in a lower energy consumption in all those environment where the limited energy is a critical property to take into account. In future works, the relationship between p and q parameters could be further investigated, running more simulations with the value of Q2PR dynamically changing with respect to network conditions and transmission characteristics. This could lead to improvements in terms of energy saving and detection effectiveness.

REFERENCES

- Basurra, S. S., M. D. Vos, J. Padget, Y. Ji, T. Lewis, & S. Armour (2015). Energy efficient zone based routing protocol for MANETs. *Ad Hoc Networks 25, Part A*, 16–37.
- De Rango, F. (2009). Trust-based SAODV protocol with intrusion detection, trust management and incentive cooperation in MANETs. *International Journal of Interdisciplinary Telecommunications and Networking (IJITN) 1(4)*, 54–70.
- De Rango, F., S. Amelio, & P. Fazio (2013, July). Enhancements of epidemic routing in delay tolerant networks from an energy perspective. In *Wireless Communications and Mobile Computing Conference (IWCMC), 2013 9th International*, pp. 731–735.
- De Rango, F., P. Fazio, F. Scarcello, & F. Conte (2014, Oct). A new distributed application and network layer protocol for VoIP in mobile ad hoc networks. *IEEE Transactions on Mobile Computing 13(10)*, 2185–2198.
- De Rango, F., F. Guerriero, & P. Fazio (2012, April). Link-stability and energy aware routing protocol in distributed wireless networks. *IEEE Transactions on Parallel and Distributed Systems 23(4)*, 713–726.
- De Rango, F. & A. Marano (2009). Trust-based SAODV protocol with intrusion detection and incentive cooperation in MANET. pp. 1443–1448.
- Fazio, P., F. De Rango, & C. Sottile (2015). A predictive cross-layered interference management in a multichannel MAC with reactive routing in VANET. *IEEE Transactions on Mobile Computing PP (99)*.
- Fazio, P., F. De Rango, C. Sottile, & A. F. Santamaria (2013). Routing optimization in vehicular networks: A new approach based on multiobjective metrics and minimum spanning tree. *International Journal of Distributed Sensor Networks 2013*.
- Fazio, P. & M. Tropea (2012). A new markovian prediction scheme for resource reservations in wireless networks with mobile hosts. *Advances in Electrical and Electronic Engineering 10(4)*, 204.
- Feeney, L. M. (2001, June). An energy consumption model for performance analysis of routing protocols for mobile ad hoc networks. *Mob. Netw. Appl. 6(3)*, 239–249.
- Hernández-Orallo, E., M. Olmos, J.-C. Cano, C. Calafate, & P. Manzoni (2014). A fast model for evaluating the detection of selfish nodes using a collaborative approach in MANETs. *Wireless Personal Communications 74(3)*, 1099–1116.
- Hidoussi, F., H. Toral-Cruz, D. E. Boubiche, K. Lakhtaria, A. Mihovska, & M. Voznak (2015). Centralized IDS based on misuse detection for cluster-based wireless sensors networks. *Wireless Personal Communications 85(1)*, 207–224.
- Kurmis, M., D. Dzemydiene, A. Andziulis, M. Voznak, S. Jakovlev, Z. Lukosius, & G. Gricius (2014). Prediction based context data dissemination and storage model for cooperative vehicular networks. In *Nostradamus 2014: Prediction, Modeling and Analysis of Complex Systems*, pp. 21–30. Springer.
- Lupia, A. & F. De Rango (2014, July). Performance evaluation of secure AODV with trust management under an energy aware perspective. In *Performance Evaluation of Computer and Telecommunication Systems (SPECTS 2014), International Symposium on*, pp. 599–606.
- Lupia, A. & F. De Rango (2015). Evaluation of the energy consumption introduced by a trust management scheme on mobile ad-hoc networks. *Journal of Networks 10(4)*.
- Ravi, G. & K. Kashwan (2015). A new routing protocol for energy efficient mobile applications for ad hoc networks. *Computers & Electrical Engineering*, –.
- Safarik, J., M. Voznak, F. Rezaca, P. Partilaa, & K. Tomalaa (2013). Automatic analysis of attack data from distributed honeypot network. In *Proc. of SPIE Vol*, Volume 8755, pp. 875512–1.
- Sun, Y., W. Yu, Z. Han, & K. Liu (2006, Feb). Information theoretic framework of trust modeling and evaluation for ad hoc networks. *Selected Areas in Communications, IEEE Journal on 24(2)*, 305–317.
- Tan, S., X. Li, & Q. Dong (2015). Trust based routing mechanism for securing OSLR-based MANET. *Ad Hoc Networks 30*, 84–98.
- Voznak, M. & J. Safarik (2012). DoS attacks targeting SIP server and improvements of robustness. *International Journal of Mathematics and Computers in Simulation 6(1)*, 177–184.

Anemia types prediction based on data mining classification algorithms

Manal Abdullah

*Department of Computer Science, Faculty of Computing and Information Technology,
King Abdul-Aziz University, Jeddah, Saudi Arabia*

Salma Al-Asmari

Department of Computer Science, Faculty of Computing, King Khalid University, Abha, Saudi Arabia

ABSTRACT: Medical Data Mining domain concerned with prediction knowledge as a method to extract desired outcomes from data for specific purposes. Anemia is one of the most common hematological diseases and in this study concentrate on the most five common types of anemia. This paper specifies the anemia type for the anemic patients through a predictive model conducting some data mining classification algorithms. The real data of dataset constructed from the Complete Blood Count (CBC) test results of the patients. These data filtered and eliminated undesirable variables, then implemented on some classification algorithms such as Naïve Bayes, Multilayer Perception, J48 and SMO using WEKA data-mining tool. Several experiments has proven that J48 decision tree algorithm gives the best potential classification of anemia types. WEKA experimenter proves J48 decision tree algorithm has the best performance with accuracy, precision, recall, True Positive rate, False Positive rate and F-measure.

Keywords: Anemia, Medical Data mining, classification algorithms, naïve Bayes, J48 decision tree, Support vector machine, SMO

1 INTRODUCTION

Data mining concept is sorting the data to identify patterns and find relationships between these data. It is techniques are appropriate for simple or structured datasets such as relational databases, trans-accional databases. Different approaches of data mining proposed to improve the challenges of storing and processing all types of data (Kaur et al., 2015 & Kishore et al., 2015).

Data mining has three basic mechanisms Clustering (Classification), Decision Rules and Analysis. Classification analyzes a set of data and produces a set of decision rules, which used to classify the data sets. In the artificial intelligence, machine learning or database systems data mining process is starting by extract the information from dataset then convert it to meaning full structure. This means that it determines patterns in datasets and embracing methods. There are many classes in data mining where the most common one is classification, which is used to predict set of relationship between data. In healthcare, it is significant to invest the development in computer technology to enhance processing the medical data such as data mining classification algorithms and tools. This paper will utilize the WEKA tool for data mining (Shouval et al., 2014). As data mining tool, WEKA name is derived from Waikato Environment

for Knowledge Analysis. It is an open source data-mining tool that provides an efficient framework for implementing several classification algorithms. This tool provides processing the datasets and filtering out and remove irrelevant (not useful) data and the dataset can be incision into test and training sets. It supports perform classification algorithms then transforming all the dataset into appropriate pattern as a machine learning form. WEKA also can upload different file formats such as ARFF, CVS, C4.5 and different databases Garner (1995).

There are growing researches interest in using data mining in the medical domain. Developing in this new approach, called medical data mining, concerned with developing systems that determine and predict knowledge from data generating from medical environments. The data mining in the medical domain specifically the hospital database, including the data, which is huge in amounts, complex in contents, with heterogeneous types, hierarchical and varying in quality. Among last years, the information on laboratories keeps on enhancing and developing. The specific patterns of information can predicated through using data mining methodologies to enhance conducting researches and evaluation of reports. The data mining classification depends on similarities existing in the data. The classification algorithms used to prove

the results is acceptable to the doctors or the end user. Medical data mining uses many algorithms such as Decision Trees, Neural Networks, Naive Bayes and others.

This paper identifies set of attributes associated with the patient CBC test result that give the anemia type, and improve the quality of prediction by identifying the anemic patients, so that can help doctors immediately improving their performance. This paper investigates the accuracy of some classification algorithms in predicting some anemia types. It is also utilizing WEKA tool for conducting classification, decision rules and analyzing the results. The evaluation of data using classification algorithms takes a set of classified data as training set and use it for training the algorithms. Then classifies the test data based on the decision rules extracted from the training set for predicting anemia diseases. The use of WEKA Experimenter conducted to specify which classification algorithm gives best performance in terms of accuracy, precision, recall, True Positive rate, False Positive rate and F-measure. The main objectives of this work are: using predictive attributes for producing data and performing data mining algorithms to get the best prediction of the anemia types using the patient Complete Blood Count (CBC) data results.

2 RELATED WORKS

There are many works that used different data mining algorithms to classify several types of diseases, such as anemia disease for specific types based on Data Mining algorithms Elshami & Alhalees (2012). In addition, many other researchers tried to find their own method. A person with anemia probably unaware of the problem because symptoms may not appear. Millions of people may have anemia and their health exposed risk. Therefore the disease is significant, several studies carried out in this domain mentioned in the literature (Yilmaz et al., 2013). (Sanap et al., 2011) developed a system using the classification technique: C4.5 decision tree algorithm and SMO support vector machine WEKA. They implemented a number of experiments using these algorithms. The anemia classification using decision tree that given clear results depend on CBC reports. (Amin et al., 2015) have compared between naive Bayes, J48 classifier and neural network classification algorithms using WEKA and working on hematological data to specify what the best and appropriate algorithm. The proposed model can predict hematological data and the results showed that the best algorithm is J48 classifier with high accuracy and naive Bayes is the lowest average in average errors. The study

of (Sanap et al., 2011) and (Amin et al., 2015) proved that the C4.5 algorithm (as J48 in WEKA) results gives high accuracy more than other classifiers. Dogan & Turkoglu (2008) based on the biochemistry blood parameters they designed a system to help physicians in the diagnosis of Anemia. The system designed using the decision tree algorithm. The system used the characteristics of the hematology and classify the results into positive or negative Anemia. The results of this system accorded with physicians' decision. Siadaty & Knaus (2006) selected decision trees as a common and simple classifier and also has low computational complexity. The problem was the needed time to build a decision tree for large dataset is come to be intractable. They solved the problem by developing a parallel model of ID3 algorithm. It is a thread-level parallelism decision tree and do the computations independently. The experiment done on anemic patient's data set. (Kishore et al., 2015) presented set of the basic classification algorithms, which group of essential types of classification methods such as decision trees, Bayesian networks, k-nearest neighbor and support vector machine classifier. The study shows a comprehensive review of diverse classification algorithms in data mining. This research presents an investigation for five types of anemia disease by using naive Bayes, Multilayer perception, J48 decision tree and support vector machine data mining algorithms depending on CBC data. The best one of classification algorithms depends on specifically in the problem domain Kesavaraj & Sukumaran (2013).

3 ANEMIA CLASSIFICATION

3.1 *What is anemia?*

It is a medical condition indicates to the reduction of hemoglobin or red cell concentration in the human blood. A Complete Blood Cell (CBC) count test conducted for patients in laboratory. The anemia disease types identified using this information: age, gender, hemoglobin, Hematocrit and other attribute values when it is lower a normal range Green (2012). Anemia types classification according to CBC test values illustrated in Fig. 1 (Sanap et al., 2011).

3.2 *The anemia classification*

Anemia disease categorized into different types based on the CBC test values. In this model Anemia types nomenclature illustrated (see Table 1) and classified according to MCV (Mean corpuscular volume) value into the three essential kinds of microcytic ($MCV < 80$) ft, normocytic

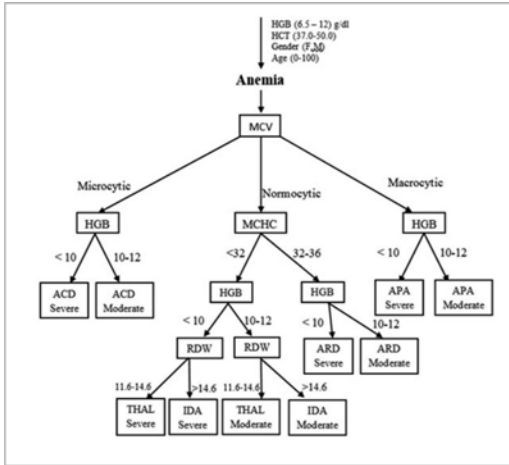


Figure 1. Anemia types classification.

Table 1. ANEMIA types nomenclature.

ACD	Anemia of chronic disease
IDA	Iron deficiency anemia
ARD	Anemia of renal disease
THAL	Thalassemia
APA	Aplastic Anemia

(MCV = 80–100) ft, and macrocytic (MCV>100) ft anemia, and classified using MCHC (Mean corpuscular hemoglobin concentration) into normochromic (MCHC = 32–36) g/dl and hypochromic (MCHC<32) g/dl anemia. RDW (Red Cell Distribution Width) used to measure the anemia and it is high if (RDW>14.6) and normal if (RDW = 11.6–14.6) Green, (2012) and (Sanap et al., 2011).

4 THE PROPOSEDMETHOD

4.1 Experimental setup

In the context of classification the anemia types. A number of attributes are considered to predict the type of anemia for the anemic patient. These influencing attributes are categorized as an input. The data is taken from Complete Blood Count (CBC) test results, which are conducted by collecting blood samples from 41 anemic patients (41 instances) and constructing ANEMIA dataset. The dataset consists of 7 attributes and defined in Table 2 along with their values.

Then data is transformed into a standard file format. CSV, which is supported by the WEKA tool to construct ANEMIA dataset, filtered and eliminating out irrelevant data using specific

Table 2. Attributes of ANEMIA dataset.

Attribute	Attribute value	Attribute category
Age	0–12 >12	Child Adult
Gender	Female Male	F M
MCV	<80 80–100 >100	Microcytic Normocytic Macrocytic
HCT	<37 37.0–50.0	Low Normal
HGB	<10 10–12	Severe Moderate
MCHC	<32 32–36	hypochromic normochromic
RDW	>14.6 11.6–14.6	High Normal

techniques. The CBC data contain 34 irrelevant attributes that are removed. The relevant attributes are shown in Table 2. The attributes are verities between nominal and numeric values and each has its own determined category.

The classification algorithms performed for predicting and classifying five most common Anemia types based on rules that shown in Table 3. The analysis of identifying anemia types are conducted using the WEKA tool. (Siadaty et al., 2006, Sanap et al., 2011 and Shashidhara, 2012):

The implementation of the proposed method starts by collecting CBC results and build our own dataset. Then data are preprocessed to extract and filter the attributes of importance. Data are converted to CSV format to be able using by WEKA classifier software. CSV file format is selected to allows data to be saved in a table structured (spreadsheet) format. After the classification and generated results, evaluated using the WEKA experimenter and the Knowledge Flow Model.

4.2 The proposed algorithms for classification

In this method, various data mining algorithms are used for predicting the anemia type for patients. During this study, classification algorithms used for prediction and the dataset are tested then analyzed with four candidate algorithms which are: Naïve Bayes, neural network (multilayer perception), Decision Tree (J48) and Support Vector Machine (SMO). The Naïve Bayes algorithm implements the principle of conditional probabilities that computes a probability by calculating the rate of values and combinations of values in the specific data. This algorithm determines the probability of an event happen given the probability of another event that has already happened. Naïve Bayes algorithm use

Table 3. Anemia classification rules.

The rule	Decision*
IF (MCV = microcytic AND HGB = 10–12) then	ACD, moderate
Else if (MCV = microcytic AND HGB = <10) then	ACD, severe
Else if (MCV = normocytic AND MCHC <32 AND RDW = 11.6–14.6 AND HGB = 10–12) then	THAL, moderate
Else if (MCV = normocytic AND MCHC <32 AND RDW = 11.6–14.6 AND HGB = <10) then	THAL, severe
Else if (MCV = normocytic AND MCHC <32 AND RDW = 11.6–14.6 AND HGB = 10–12) then	IDA, moderate
Else if (MCV = normocytic AND MCHC <32 AND RDW = 11.6–14.6 AND HGB = <10) then	IDA, severe
Else if (MCV = normocytic AND MCHC = 32–36 AND HGB = 10–12) then	ARD, moderate
Else if (MCV = normocytic AND MCHC = 32–36 AND HGB = <10) then	ARD, severe
Else if (MCV = macrocytic AND HGB = 10–12) then	APA, moderate
Else if (MCV = macrocytic AND HGB = <10)	APA, severe

*The decision includes (Anemia type and severity grade).

kernel density estimators that improve implementation if the normal assumption clearly correct; it can also deal with numeric attributes using supervised discretization Vijayarani & Muthulakshmi (2013). The second algorithm is a neural network in WEKA named (multilayer perception). It is a feed forward neural network multilayer model that can map set of the input data (each one is a neuron) into a set of suitable outputs. The input node is an element with a nonlinear activation function. The multilayer perception consists of multiple one or more of hidden layers of nodes called (hidden neurons) in a directed chart, with each layer completely connected to the next layer (Prakash et al., 2015). The J48 decision tree algorithm is used also for automatic processing and canchoose related aspects from training data. It can cut the meaningless approaches into effective process, especially when dealing with continuous attributes. It split the values based on the thresholding to specify what is upper than, less than or equal to the threshold value. J48 algorithm contains the capability of dealing with training data with missing values of some attributes (Ahmad et al., 2011). Support Vector Machines named (SMO) in WEKA used as a supervised learning method which analyzing data and recognizing patterns. It is not probable classifier, which process set of input data and

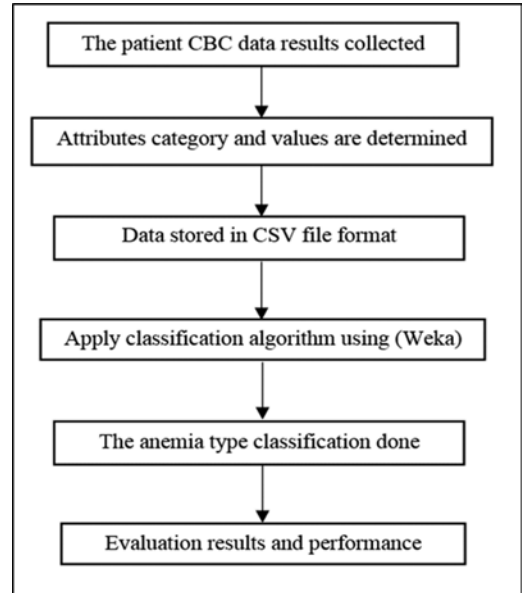


Figure 2. Flowchart of proposed method.

classify it into what is the two probable classes and give the output. The SVM algorithm has the same functional form of neural networks and radial basis functions Kesavaraj & Sukumaran (2013). It is generally used to a two class classification problem, its detect the plane and gives the greatest separation between the two classes. The SVM algorithm discovers the optimal plane with a maximum distance to the nearby point of the two classes. A set of instances that are closest to the optimal plane, explains the support vector and specify the margins of each class (Shouval et al., 2014). See the description of the proposed methodology illustrated as flowchart shown in Fig. 2.

5 RESULTS AND DISCUSSION

Evaluation of data is done by using 41 instances in the dataset using Naïve Bayes, neural network in WEKA (multilayer perception), J48 decision tree algorithms, and support vector machine in WEKA (SMO) with the test option: several percentages splits (20%, 40%, 60%) of the dataset see Table 4.

The results in Table 4 of evaluation an ANEMIA dataset using WEKA through different experiments 20%, 40%, 60% percentage split data. The table include the result through accuracy (correctly classified instances), mean absolute error, weighted average ROC and F-measure. Fig. 3 show the SMO algorithm results using 60% training set data.

Table 4. Simulation result of algorithms using 20%, 40%, 60% training set data.

Algorithm	Training Set	Accuracy*%0	Mean absolute error%0	Weighted av. ROC	F-Measure
Naïve Bayes	20%	30.303	0.458	0.507	0.257
	40%	60	0.3372	0.708	0.587
	60%	68.75	0.2645	0.825	0.68
Multilayer Perception	20%	39.3939	0.3744	0.775	0.383
	40%	72	0.2198	0.852	0.716
	60%	87.5	0.1372	0.921	0.859
J48 Decision tree	20%	27.2727	0.3207	0.855	0.218
	40%	88	0.1689	0.868	0.878
	60%	93.75	0.1743	0.97	0.935
SMO	20%	39.3939	0.4108	0.677	0.396
	40%	84	0.2578	0.902	0.83
	60%	93.75	0.2361	0.96	0.912

*Correctly Classified Instances.

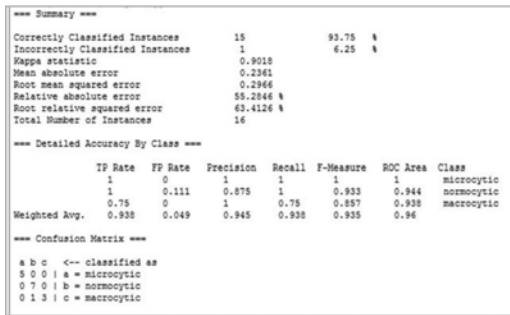


Figure 3. Support vector machine (SMO) algorithm output using 60% training set data.

The test that using percentage split is conducted by deciding a specific percent of data for training and the rest of data for testing. In this experiment the percentage split are chosen as 20%, 40% and 60%, where the partitions is conducted randomly. The percentage split 20%: the data will split into 20% will used as training set data and the rest 80% will used as testing set data. The same process done with other percentages 40% and 60%.

The accuracy (Correctly Classified Instances) rate of the results using different splitting percentages increased in naïve Bayes, J48, multilayer perception and SMO. The accuracy increasing with the training set average respectively. All statistic results provide an important comparison of the accuracy between all algorithms done and finally it have been investigated that J48 decision tree and SMO algorithms implement best results with accuracy 93.75% when using the percentage split 60%. The accuracy measure of all the algorithms using 60% training set are illustrated in Fig. 4.

The results shown in the Table 5 are the performance of naïve Bayes, neural network (multilayer perception), J48 decision tree and SMO using

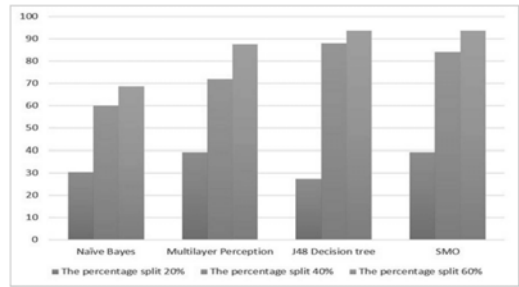


Figure 4. Comparing algorithms accuracy using the percentage split 60%.

WEKA experimenter. The data mining measures in the table illustrates more useful and precise evaluation of algorithm’s performance, especially when dealing with datasets: recall (sensitivity), precision, F-measures, true positive rate and false positive rate, which computed as follows:

$$\text{Recall (sensitivity)} = \frac{\text{True Positive rate}}{(\text{True Positive rate} + \text{False Negative rate})}$$

$$\text{Precision} = \frac{\text{True Positive rate}}{(\text{True Positive rate} + \text{False Positive rate})}$$

$$\text{F-measure} = \frac{(2 * \text{recall} * \text{precision})}{(\text{recall} + \text{precision})}$$

The True Positive rate is the number of positive instances classified correctly, The False Negative rate is the number of positive instances (records) classified negatively; False Positive rate is the number of negative instances classified positively (Huang et al., 2012).

In the context of using WEKA experimenter a snapshot of using F-measure illustrated in Fig. 5, using the precision in Fig. 6 and using the TP rate

Table 5. Comparison of classification algorithms.

Algorithm	TP Rate	FP Rate	Precision	F-Measure	Recall
Naïve Bayes	0.92	0.10	0.93	0.91	0.92
Multilayer Perception	0.92	0.10	0.95	0.91	0.92
J48 Decision tree	0.93	0.05	0.97	0.93	0.93
SMO	0.90	0.40	0.85	0.84	0.90

```

Tester: weka.experiment.PairedCorrectedTester
Analysing: IR_precision
Datasets: 1
Resultsets: 4
Confidence: 0.05 (two tailed)
Sorted by: -
Date: 29/12/15 01:58 p

Dataset (1) bayes.N | (2) func (3) func (4) tree
-----
ANEMIA (10) 0.93 | 0.95 0.85 0.97
-----
(v/ /*) | (0/1/0) (0/1/0) (0/1/0)

Key:
(1) bayes.NaiveBayes
(2) functions.MultilayerPerceptron
(3) functions.SMO
(4) trees.J48
    
```

Figure 6. Comparing algorithms with use the WEKA experimenter using precision.

```

Tester: weka.experiment.PairedCorrectedTester
Analysing: F_measure
Datasets: 1
Resultsets: 4
Confidence: 0.05 (two tailed)
Sorted by: -
Date: 29/12/15 01:44 p

Dataset (1) bayes.N | (2) func (3) func (4) tree
-----
ANEMIA (10) 0.91 | 0.91 0.84 0.93
-----
(v/ /*) | (0/1/0) (0/1/0) (0/1/0)

Key:
(1) bayes.NaiveBayes
(2) functions.MultilayerPerceptron
(3) functions.SMO
(4) trees.J48
    
```

Figure 5. Comparing algorithms with use the WEKA experimenter using F-measure.

```

Tester: weka.experiment.PairedCorrectedTester
Analysing: True_positive_rate
Datasets: 1
Resultsets: 4
Confidence: 0.05 (two tailed)
Sorted by: -
Date: 30/12/15 01:23 p

Dataset (1) bayes.N | (2) func (3) tree (4) func
-----
ANEMIA (10) 0.92 | 0.92 0.93 0.90
-----
(v/ /*) | (0/1/0) (0/1/0) (0/1/0)

Key:
(1) bayes.NaiveBayes
(2) functions.MultilayerPerceptron
(3) trees.J48
(4) functions.SMO
    
```

Figure 7. Comparing algorithms with use the WEKA experimenter using true positive rate.

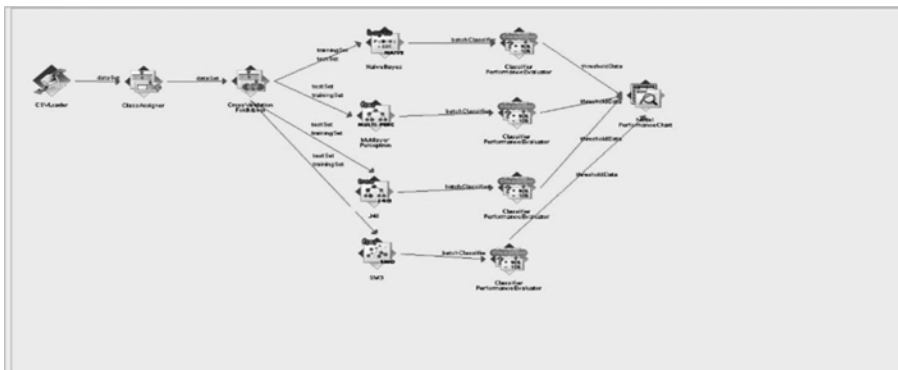


Figure 8. Knowledge flow model using WEKA.

in Fig. 7. In these experiments, it has shown that J48 decision tree performs best among four algorithms with F-Measure 93%, Sensitivity is 93%, true positive rate is 93%, Precisions 97% and it is the lowest in the false positive rate 0.05.

The comparative performance based on the accuracy among four algorithms also conducted by using knowledge flow model shown in Fig. 8,

which shows the membership tree structure using 10 folds validation test.

The performance chart of knowledge flow model conducted for the experiment algorithms Naive Bayes, Multilayer Perceptron, J48 and SMO. It is another important performance measures in WEKA. The performance represented by the Region of meeting Curve (ROC) for each

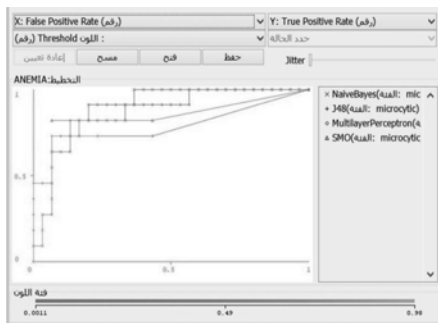


Figure 9. Performance chart of (ROC) curve.

algorithm based on 10 folds validation test. From the Fig. 9, it is clearly shown that J48 decision tree has the highest weighted average ROC0.97.

6 CONCLUSION AND FUTURE WORK

This paper used many classification algorithms to get the best prediction of Anemia types based on a dataset of 41 patients. The proposed model is designed depending on five most common anemia types then classifying and analyzing the anemia type for anemic patients' dataset.

The dataset constructed from results of complete blood count test CBC. The experiment conducted by using four data mining classification algorithms where J48 decision tree and SMO performs best with 93.75% accuracy in the percentage split 60%.

When comparing the selected algorithms through utilizing of WEKA experimenter is proved that the J48 decision tree algorithm gives the best performance with F-Measure, Sensitivity, The true positive rate, Precisions and the lowest value in the false positive rate. Therefore, J48 proved to be potentially the most effective and efficient classification algorithm. In the same context, based on anemia model the performance chart by Region under meeting Curve (ROC) shown that the highest weight for J48 decision tree.

In future, use more of the data mining algorithms to classify all types of anemia diseases on different datasets to find the accuracy and predictions of preferred results.

REFERENCES

Ahmad, A., Mustapha, A., Zahadi, E. D., Masah, N., & Yahaya, N. Y. (2011). Comparison between Neural Networks against Decision Tree in Improving Prediction Accuracy for Diabetes Mellitus Digital Information Processing and Communications (pp. 537–545): Springer.

Amin, M. N., & Habib, M. A. Comparison of Different Classification Techniques Using WEKA for Hematological Data.

Dogan, S., & Turkoglu, I. (2008). Iron-deficiency anemia detection from hematology parameters by using decision trees. *International Journal of Science & Technology*, 3(1), 85–92.

Elshami, E. H., & Alhalees, A. M. (2012). Automated Diagnosis of Thalassemia Based on Data Mining Classifiers. Paper presented at the The International Conference on Informatics and Applications (ICIA2012).

Garner, S. R. (1995). Weka: The waikato environment for knowledge analysis. Paper presented at the Proceedings of the New Zealand computer science research students conference.

Green, R. (2012). Anemias beyond B12 and iron deficiency: the buzz about other B's, elementary, and nonelementary problems. *ASH Education Program Book*, 2012(1), 492–498.

Huang, F., Wang, S., & Chan, C.-C. (2012). Predicting disease by using data mining based on healthcare information system. Paper presented at the Granular Computing (GrC), 2012 IEEE International Conference on.

Kaur, P., Singh, M., & Josan, G. S. (2015). Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector. *Procedia Computer Science*, 57, 500–508.

Kesavaraj, G., & Sukumaran, S. (2013). A study on classification techniques in data mining. Paper presented at the Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on.

Kishore, C. R., Rao, K. P., & Murthy, G. Performance Evaluation of Entorpy and Gini using Threaded and Non Threaded ID3 on Anaemia Dataset. *Life*, 6(10), 10–12.

Prakash, V. A., Ashoka, D., & Aradya, V. M. (2015). Application of Data Mining Techniques for Defect Detection and Classification. Paper presented at the Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014.

Sanap, S. A., Nagori, M., & Kshirsagar, V. (2011). Classification of anemia using data mining techniques Swarm, Evolutionary, and Memetic Computing (pp. 113–121): Springer.

Shashidhara, M. Classification of Women Health Disease (Fibroid) Using Decision Tree algorithm.

Shouval, R., Bondi, O., Mishan, H., Shimoni, A., Unger, R., & Nagler, A. (2014). Application of machine learning algorithms for clinical predictive modeling: a data-mining approach in SCT. *Bone marrow transplantation*, 49(3), 332–337.

Siadat, M. S., & Knaus, W. A. (2006). Locating previously unknown patterns in data-mining results: a dual data-and knowledge-mining method. *BMC Medical Informatics and Decision Making*, 6(1), 13.

Vijayarani, S., & Muthulakshmi, M. (2013). Comparative Analysis of Bayes and Lazy Classification Algorithms. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(8), 3118–3124.

Yilmaz, A., Dagli, M., & Allahverdi, N. (2013). A fuzzy expert system design for iron deficiency anemia. Paper presented at the Application of Information and Communication Technologies (AICT), 2013 7th International Conference on.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Energy Efficient Optimized Routing Algorithm (EEORA)

N. Alharbe

College of Computer Science and Engineering, Taibah University, Al Madinah al Munawwarah, Saudi Arabia

M. Abdullah

Faculty of Computing and Information Technology FCIT, King Abdulaziz University KAU, Jeddah, Saudi Arabia

ABSTRACT: As a part from the great revolution that was occurred recently, a continuous improvement was and still achieved in communication and network field. Wireless Sensor Networks was emerged as a result for this development, extensive researches were introduced in order to improve its performance since it is now employed within several tactical applications, including; detection and monitoring applications. Routing process is considered one of the hottest research areas throughout the last years, several routing protocols were proposed in order to improve the performance of the network. In this paper, a new Energy Efficient Optimized Routing Algorithm (EEORA) is proposed and implemented by hybridization for the features of flat and Hierarchical routing techniques. EEORA routing process is performed into two stages, clustering and data communication phases. MATLAB program was employed in simulating the model of the proposed routing protocol for both Wireless Sensor Networks (WSNs) and Dense Wireless Sensor Networks (DWSNs). The results showed that more scalability is obtained by EEORA over other considered routing protocols. The results also showed that improved performance in terms of, energy consumption, end-to-end delay, network life time and throughput is also achieved by EEORA.

1 INTRODUCTION

1.1 Background

Recently, a rapid growing has been achieved in “Wireless Sensor Networks (WSNs)” and it is considered one of the hottest research areas due to its wide range of applications in several practical life fields, including, habitat monitoring, military surveillance, environmental observation, etc. WSN area consists of several sensing nodes prepared with sensing, processing ability, data gathering, limited energy capability and storage space. Sensor nodes are provided with a platform by the WSN for monitoring and sensing the concerned environmental area (Ali and Parmanand, 2015 and Srinath et al, 2007). Due to its importance, WSN is an emerging technology, which will achieve influential impact on humanity lives future. WSN is extensively studied by several researchers in order to overcome its constraints and complexity, including; communication bandwidth, network density, power source, data aggregation, memory, processor and mobility (Al-Fares et al, 2009).

In WSN, the efficiency of the used protocol for routing plays an essential and significant role in the transition of the data packet inside the network. The conventional protocols of routing give a little consideration for node energy consumption. Since

the nodes energy is limited, the maximum WSN lifetime is a significant goal to the routing protocol destination. Therefore, it is essential to consider the energy balance consumption in addition to the consumption of energy along the path needed to transmit small messages toward the required destination. Moreover, since there is a large number of sensor nodes in these networks, especially in “Dense WSNs (DWSNs)”, only partial information regarding to the topology information can be obtained by the nodes, so it is essential for the routing protocol to select the optimal path based on these available partial information (Guo and Tang, 2010).

1.2 Related works

Several researchers throughout the last years have extensively studied routing process in WSN and DWSNs. During their work, they proposed several protocols that can be employed in routing the information between network entities. Some of these protocols are now introduced. Routing protocols in WSNs are classified into two main groups, which are; flat-based and hierarchical-based routing protocols. In flat-based type, all nodes inside the network perform similar tasks. Each node is considered as potential BS. All nodes are assigned unique identifier in order to be used as node identity representative. The most common examples

for flat-based routing protocols are; “Directed Diffusion (DD)” and “Sensor Protocols for Information via Negotiation (SPIN)”. Hierarchical-based type is also known as cluster-based routing protocols; the total area is divided into number of clusters and a CH is selected within each cluster according to some criterion as a representative for all node within correspondence cluster. The normal nodes send their data to CH but it is not allowed for nodes to communicate with BS directly. CH in turns aggregates the data from several nodes and transmits it to BS. Common examples for this type are; “Low Energy Adaptive Clustering Hierarchy (LEACH)”, “Power-Efficient Gathering in Sensor Information Systems (PEGASIS)”, “Two-Level LEACH (TL-LEACH)”, “DIRECTED DIFFUSION LEACH (DD-LEACH)” (Geetu and Juneja, 2013, Kodali and Sarma, 2013, Singh et al, 2010, Krishnaveni and Sutha, 2012).

Khushboo and Daniel (2015) proposed a “Section Based Hybrid Routing Protocol for heterogeneous WSN using Artificial bee colony (SBHRA)”. In their model, some nodes are permitted to directly communicate with BS and the remaining nodes communicate with BS through employing a clustering method. “Cluster Heads (CHs)” have been selected employing the concept of “Artificial Bee Colony”. The results confirmed the effectiveness of their hybrid model in minimizing the sensors consumption of energy and in improving the throughput and lifetime of the network.

A dynamically reconfigurable protocol for routing has been proposed by Gao and Piao (2014). They utilize the potential concept in physics in their design. The user was provided an access to construct diverse virtual potential fields of hybrid type; this will in turns make the proposed protocol applicable to be used for several applications. In order to improve the proposed protocol effect, a parameter was set up that can be dynamically changed in order to affect the virtual field. The routing protocol can be constantly optimized through regulating this parameter. The administrator was also provided by suitable method for the purposes of reconfiguring the protocol through remote application. The results confirmed that the proposed protocol is able to be easily reconfigured.

Zhao et al (2012) introduced an improved version of LEACH protocol aimed to reduce the consumed energy and hence improving the network lifetime. They introduced a timer in order to confirm that the optimal node has been selected as CH during the set up phase. The energy has been effectively exploited through employing a hybrid routing of multi-hop and single-hop. The results confirmed the usefulness of the proposed model in reducing the consumption of the energy and hence increasing the WSN lifetime.

Another routing protocol for WSNs was proposed by Song et al (2010). They exploit “Artificial Fish Swarm Optimization (AFSO)” in their model to form the clusters. Employing this algorithm aimed to overcome the NP-hard Problem in which optimal k clusters are formed according to specific rules. They compared the performance of their proposed model with ordinary LEACH and LEACH-C. The results confirmed the effectiveness of the proposed protocol in improving the network life and reducing the consumption of the energy.

Ya et al (2014) proposed a new routing algorithm employing the beaconing to be used within WSNs. their protocol aimed to achieve energy efficiency and reliability while transferring both information dissemination and collection for beaconing packets that are exchanged within WSNs. node-node scheme of routing is also supported by their proposed model. The simulation results confirmed the effectiveness of their protocol to be used within WSNs. they proved that the reliability can be guaranteed using this protocol. Furthermore, the protocol enhance the network lifetime by more than 20% compared to “Collection Tree Protocol (CTP)”.

Malathi et al (2012) proposed a new protocol of routing for WSNs aimed mainly to improve the lifetime of WSNs. They proposed a new clustering method was proposed so that the energy consumption can be reduced and the distributed equally between nodes in the WSNs. This will in turns results in enhancing the lifetime of WSNs. The sensor field is divided into several clusters and a node is then elected as CH, the routing tree is then constructed to forward the data toward BS. All nodes use only single hop to transfer their data to CH. The aggregated data is then forwarded from CH to BS. The results proved that the proposed protocol enhanced the performance by 31%, 18% and 15% when compared to LEACH, LEACH-C and TL-LEACH respectively.

Li et al (2013) stated that the existing routing protocols are not able to ensure the optimal section of CHs due to the comprehensive consideration lack about the nodes’ remaining energy. They also dedicated that the communication of single-hop type through CHs may causes an imbalance within CHs energy. During their work, they hybridize an improved version from the algorithms of “Particle Swarm” and inter-clustering routing in order to formulate “adaptive Energy-Efficient Clustering Routing Protocol (AECRP)”. The results proved the protocol effectiveness in terms of data reliability, energy consumption and improved network lifetime.

A novel protocol called “Density controlled Divide-and-Rule (DDR)” was proposed by Ahmad et al (2013) to be used especially for

DWSN. DDR aimed mainly to overcome the energy hole and coverage hole problems. Their model considered uniform nodes distribution with fixed CHs number throughout each round. They adopted the technique of static clustering during the formation of the clusters. The segmentation of the network area is performed in a way results in reducing the distance between the nodes and CHs and from CHs to BS. They divided the network area into concentric squares and then divided the area that separates each two concentric squares into four equally area segments. Each segment contains a number of nodes. The CHs are dynamically changed each round, the nearest CH to BS is selected as CH in first round and second least node during the second round, this process is continuously performed throughout the communication process. The simulation results proved that DDR is better than LEACH in terms of the sent packets to BS and hence reduced the consumed energy so that the lifetime of network is improved.

An Improved version from DDR known as (IDDR) was proposed by Saleem et al (2014). They utilized the concept of uniform consumption of the energy. They followed same cluster formation as used in DDR, but the CHs are selected based on the maximum amount of residual energy. They compared IDDR with DDR and the result confirmed that IDDR outperforms DDR in terms of different performance criteria.

2 FORMULATION OF THE PROBLEM

2.1 Protocol flow chart

As stated early, the goal from this paper is to design and model hybrid routing protocol to be used within DWSN. The new designed protocol will be names as “Energy Efficient Optimized Routing Algorithm (EEORA)” and its mainly aims to reduce the amount of consumed energy within the DWSN so that the energy can be efficiently utilized sand to efficiently transmit the data through the network. This will be achieved via reducing two main metrics, which are; the distance required to discover, maintain and use the routes and the message sent amount within the network. The proposed routing protocol utilizes the benefits of flat-based and hierarchal-based routing protocol, so the concepts of both types of routing protocols will be combined together. The main idea of EEORA is to employ a new clustering algorithm in splitting the DWSN into specific number of clusters and to utilize the concept of flat-based routing while applying new approach in data communication stage that mainly based on minimizing the number of sensor nodes that transmit the data.

Two main stages are included within EEORA in order to perform the routing process, which are; formation of clusters and data communication. The flow chart of EEORA is shown in Figure 1 below.

2.2 Clustering process

A novel clustering algorithm will be followed in order to divide the DWSN into several clusters. The clustering process is initiated by dividing the whole area of DWSN into specific number of concentric squares. This division is performed so that the distance between the normal nodes and CH and also between the CH and the BS are reduced. The concentric squares formation is performed in reference with the coordinates of the BS that is assumed to be located at the center of the DWSN with coordinates (x_0, y_0) . The concentric squares number (k) is calculated using an equation shown in the previous flow chart and depends mainly on the area of the DWSN, the coordinates of the BS and the distance between the two successive concentric squares. The k^{th} square dimensions can be found using the following equations;

$$T_R^{S_k}(x_1, y_1) = (x_0 + r_m, y_0 + r_m) \quad (1)$$

$$B_R^{S_k}(x_2, y_2) = (x_0 + r_m, y_0 - r_m) \quad (2)$$

$$T_L^{S_k}(x_3, y_3) = (x_0 - r_m, y_0 + r_m) \quad (3)$$

$$B_L^{S_k}(x_4, y_4) = (x_0 - r_m, y_0 - r_m) \quad (4)$$

where;

$T_R^{S_k}$ (top right of k^{th} square), $B_R^{S_k}$ (bottom right of k^{th} square), $T_L^{S_k}$ (top left of k^{th} square), $B_L^{S_k}$ (bottom left of k^{th} square) and r_m is the distance between the BS and the boundary of k^{th} squares as given in equation below;

$$r_m = m \cdot r \quad (5)$$

Once the DWSN area divided into concentric squares, the area enclosed between each two neighboring squares will be split into four regions with equal area and each one of these region will form a cluster. The segmentation will performed for all two neighboring squares. A CH is then selected within each cluster based on the maximum amount of residual energy in addition to the minimum distance among CH and CHs within all directly connected levels. to finally form the clustered DWSN as shown in Figure 2 below. Only one CH is assigned to each cluster and the nodes in each cluster can connect any one of the neighboring CHs based on the minimum distance. This will in turns result in

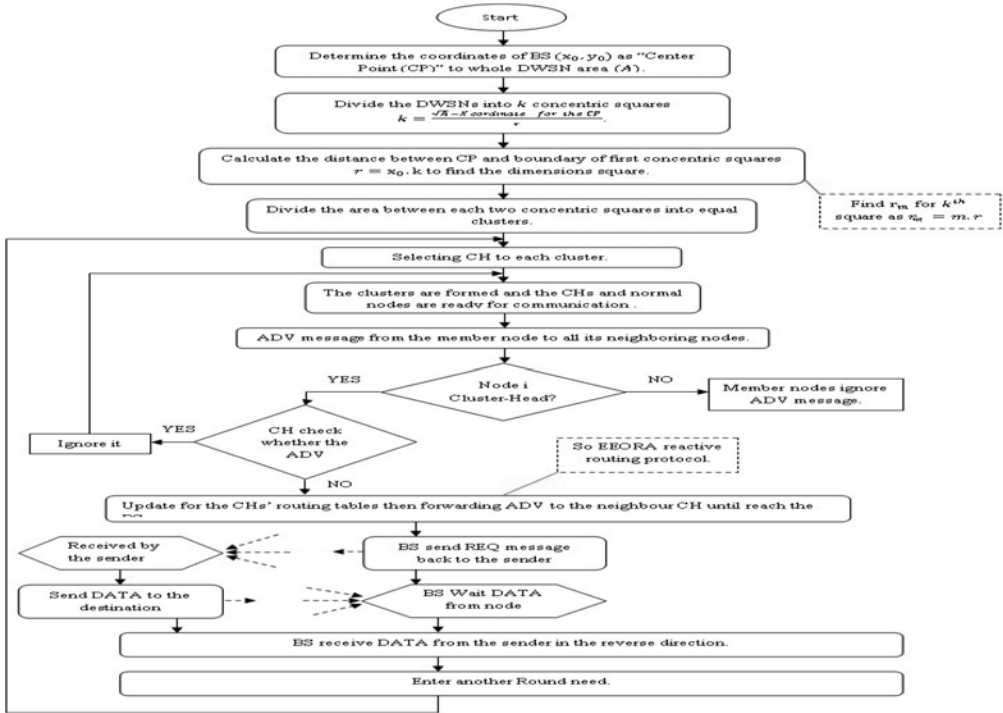


Figure 1. EEORA flow chart.

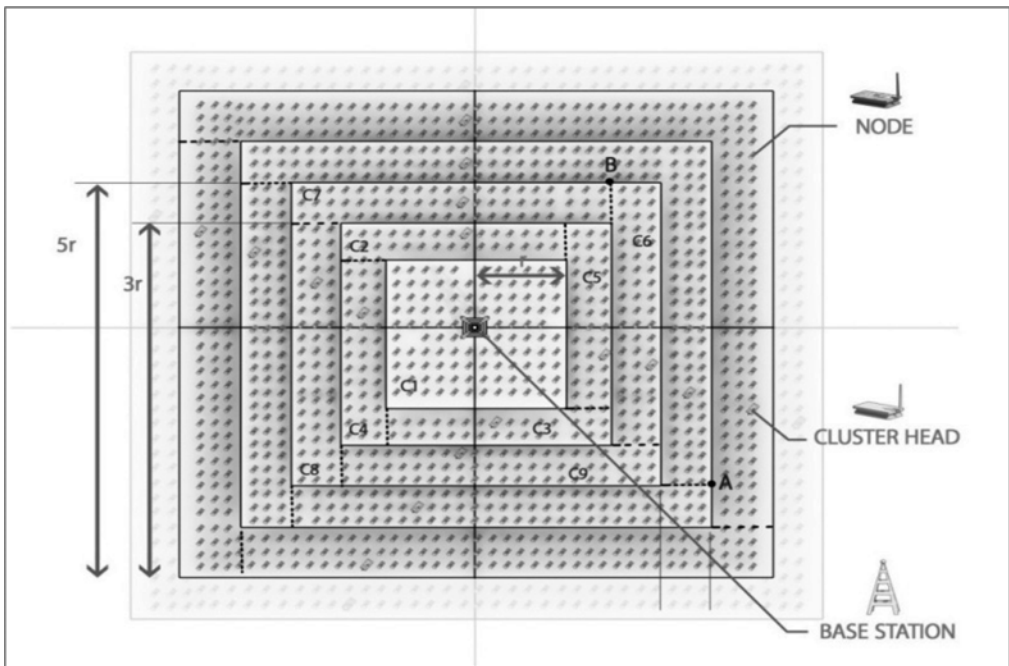


Figure 2. Final clustering approach of DWSN.

reducing the distance needed to reach the BS and hence enhance the lifetime of the DWSN.

2.3 Data communication

After the Clusters being formed and the CH being selected within each cluster, the data communication process can be initiated. Actually, novel approach for data communication is proposed here, which can be summarized in terms of three messages as listed below;

“Advertise Message (ADV)”: when a node sensed a new data and it want to share these data with the remaining nodes inside DWSN, then an ADV message is firstly employed by this node to inform the network that it has a data to be shared. The ADV message includes receiver node ID, sender node ID and No. of Hop field.

“Request Message (REQ)”: once the data being shared, the node/nodes concern in this data will respond via sending REQ message so that the sender is informed that this node is intended to receive the sensed data by the node that sent the ADV message.

“Data Message (DATA)”: once the node that initiated the process of communication via sending ADV message is informed that there is a specific node wants its sensed data, then it start sharing its data by transmitting DATA message including also

the metadata. The data communication process is summarized below Figure 3.

According to Fig. 3, seven steps hierarchal steps are required to be performed so that the communication process is performed; these steps are now summarized. Firstly, the node that senses the event from the surrounding area will broadcast an ADV message to the DWSN, this message will be only considered by CH and it is ignored by normal nodes. The CH in turns will forward this message to the neighbor CHs in case that it was not previously considered and received by that CH. These steps are continuously performed by all CHs inside the DWSN until the ADV message is received by the BS. Once the BS received the ADV message, a REQ message is sent back to the sender following reverse direction. When the sender receive the REQ message, the data can now be transmitted from the source to the destination following the routing path that were determined early when ADV and REQ messages are being transferred in DWSN. Regarding to the energy consumption while transmitting and receiving messages in the network, a similar model to that adopted in [IRRD] will followed here. In this model, the energy of the nodes decreases due to the transmission of data by an amount proportional to distance between each two nodes included in communication process. Moreover, the nodes energy will also decrease due to the reception of the data.

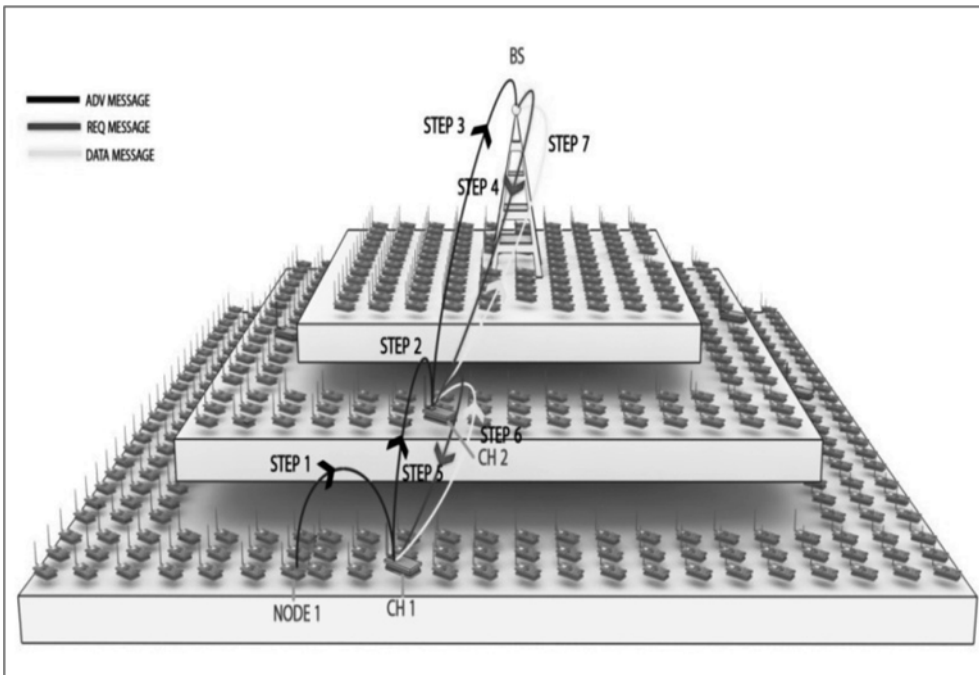


Figure 3. Data communication process.

The following equations illustrate the process of calculating the consumed energy while exchanging a packet of size L bits;

$$E_{Tx}(d, x) = E_{elec} * L + \epsilon_{amp} * L \quad (6)$$

$$E_{Rx}(L) = E_{elec} * L \quad (7)$$

ϵ_{amp} is calculated in relation to a reference distance d_0 as given below;

$$\epsilon_{amp} = \epsilon_{fs} * d^2, d \leq d_0 \quad (8)$$

$$\epsilon_{amp} = \epsilon_{fs} * d^4, d \geq d_0 \quad (9)$$

where; d denotes the distance the separates each two nodes, ϵ_{fs} and ϵ_{mp} are energy parameters correspond to the communication process. E_{elec} is the consumption of electronic energy, E_{Tx} and E_{Rx} denotes the consumed energy due to transmitting and receiving the packets respectively.

3 RESULTS AND ANALYSIS

All previously mentioned details were simulated in MATLAB simulator in order evaluate the performance of the proposed protocol and to validate its effectiveness by comparing it with different proposed routing protocols in the literature for both WSNs and DWSNs. Several metric were employed in evaluating the performance, which are; “Packet Delivery Ratio (PDR)”, Scalability, “end-to-end delay”, consumption of the energy, throughput and the life time of the DWSN. Table 1 below summarizes the simulation parameter for the proposed protocol.

Figure 4 below illustrates the simulated model of EEORA in MATLAB environment.

The previously mentioned metrics will be used in evaluating the performance of EEORA protocol. The performance of EEORA will be compared to WSNs protocols, SPINN and LEACH and it will also compared to most recent DWSNs protocols, DDR and IDDR. The first metric that will be used for performance evaluation is the consumption of the energy. The energy model that will be used in EEORA was presented early. Figure 5 below illustrates the consumed energy by these protocols after running the simulation for 20 rounds.

According to Figure 5; EEORA protocol consumes the lowest amount of energy when it is compared to WSNs or DWSNs routing protocols. Actually, this reduction in the consumed energy results from the fact that EEORA selects the optimal path and reduces the redundancy of the packets in different phases. LEACH and SPIN consumed 23% and 67.77% energy respectively

Table 1. Simulation parameters.

Parameter	Remarks
Area (A)	100 m × 100 m
Number of nodes	100, 200, 300, 400, 500
E_{elec}	50 nJ/bit
E_{da}	5 nJ/bit/signal
e_{fs}	0.001 nJ/bit/m2
e_{amp}	0.000013 nJ/bit/m4
Packet size	1000 bytes
Initial energy	2 J
Node distribution	Normal distribution
Base station location	Random
Transmission range	35 m
Number of rounds	50
Protocols	LEACH, SPIN, IDDR, DDR, EEORA
Propagation model	Distance based path loss model
Antenna type	Omni-directional
Carrier frequency	2.1 GHz
Transmit power	3 W
Miscellaneous loss	5 dB
Receiver sensitivity	-90 dBm

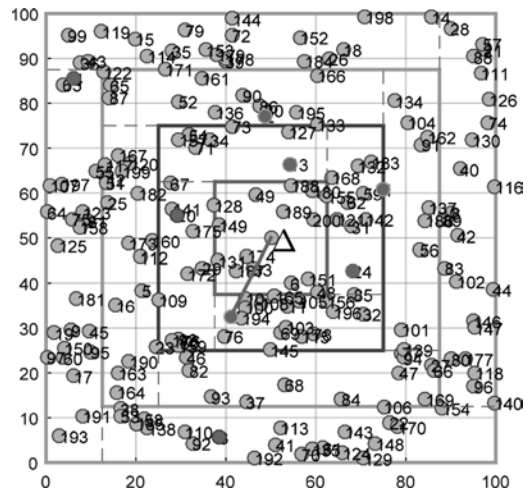


Figure 4. EEORA model in MATLAB environment.

more than EEORA. When employing LEACH as a routing protocol, the network is considered as single-hop, so there will be a considerable reduction in the energy compared to EEORA. It is also noticeable that EEORA achieves an improved performance over DWSNs protocols, DDR and IDDR, by reducing less amount of energy. The second metric that was considered to evaluate the performance is the achieved throughput as illustrated below in Figure 6.

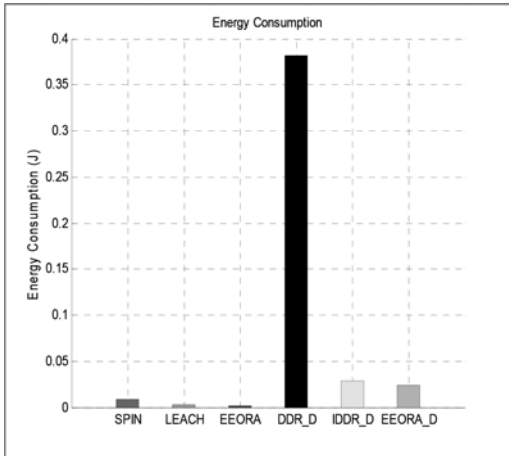


Figure 5. Consumed Energy by different routing protocols for WSNs and DWSNs.

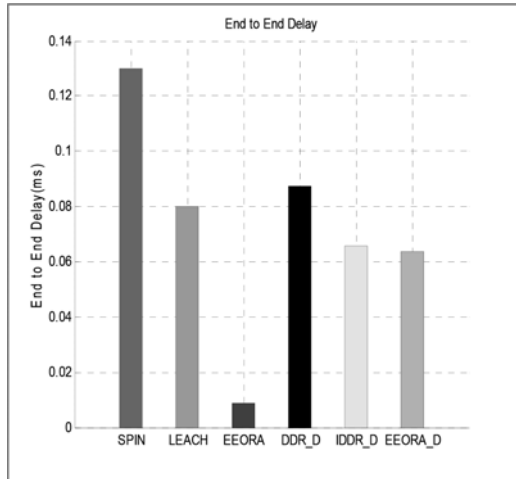


Figure 7. End-to-end delay by different routing protocols for WSNs and DWSNs.

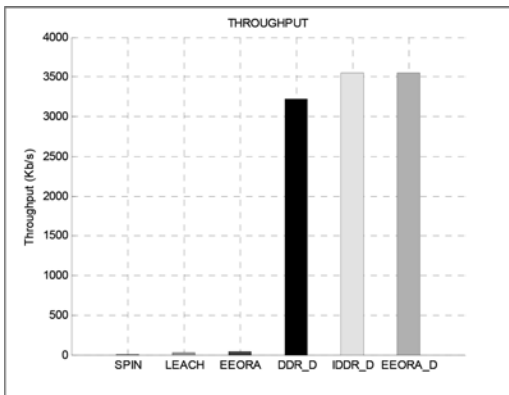


Figure 6. Achieved throughput by different routing protocols for WSNs and DWSNs.

According to Figure 6; higher throughput is achieved by EEORA compared to WSNs and DWSNs routing protocol. This enhancement is due to the fact the consumed less amount of energy, which in turns improves the lifetime of the network and hence increases the throughput. The achieved throughput for EEORA is greater by 82% and 45.6% than the achieved throughput in SPIN and LEACH respectively. Reliable paths are provided by EEORA and the data is routed efficiently to finally improve the overall network throughput. Another metric that was employed in evaluating the performance is the end-to-end delay that is recognized as the time required by the data packet to get the destination starting from the source. This delay includes; transmission, propagation, processing and queuing time. Figure 7 below illustrates the

obtained performance for the considered protocols in terms of this metric.

According to Figure 7; it is noticeable that the lowest amount of delay is achieved by EEORA in both WSNs and DWSNs. Weighted clustering method, REQ and ADV stages are implemented by EEORA in order to attain successful packet delivery. Due to the fact that the packets are transmitted to nearest CH until reach the BS, then the system end-end delay is reduced. Compared to EEORA, LEACH and SPIN achieved 87.5% and 91.67% more delay respectively. Improved delay performance is also achieved for EEORA compared to this achieved by DDR and IDDR. An essential performance metric to be considered during our evaluation is the network life time that reflects the network life span and illustrates how long the WSN can maintain. The life time of the network is closely related to the dead nodes' number, which in turns related to the speed of consuming the energy of the node. Figure 8 below illustrates life time metric for the considered protocols.

According to Figure 8; improved life time performance is achieved by EEORA over both WSNs and DWSNs routing algorithms. This enhancement is related to the reduction with the consumed energy in EEORA as shown early. for 100 nodes; the WSNs life time is around 200 hours compared to only 40 hours and 5 hours for LEACH and SPIN respectively. Moreover, in case of DWSNs, the network life time is around 73 hours compared to only 52 hours and 40 hours for DDR and IDDR respectively. Therefore, EEORA outperforms other considered routing algorithms for both WSNs and DWSNs. The performance of

EEORA was evaluated in terms of the scalability and compared to the considered routing protocols. Scalability is a significant metric to evaluate the performance of the routing protocol; it measures the network ability to handle the modification within the network. This metric is essential

to determine if increasing the network size can be handled or not. Increasing the size of the network results in more load, more loss within the packets in addition to requiring more control information which in turns results in more overhead. The following figure illustrates the scalability of the five routing protocols, while varying the network size from 100 nodes to 500 nodes, measured in terms of throughput, energy consumption and energy-delay product.

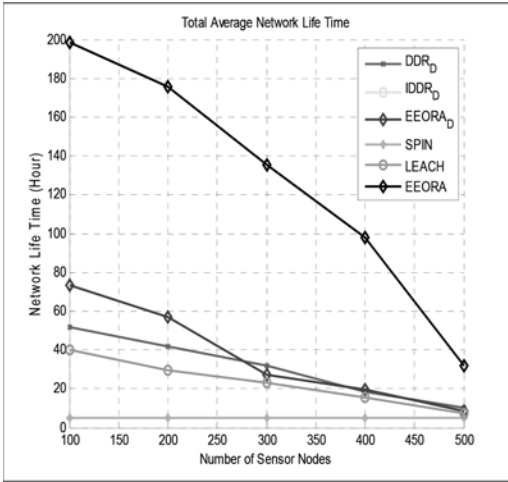


Figure 8. Network Life time for different routing protocols for WSNs and DWSNs.

According to Figure 9; increasing the nodes' number results in improving the achieved throughput when EEORA is applied within both WSNs and DWSNs. This enhancement occurred due to the increase in number of the packets successfully received by the BS. It can be also noticed that increasing the size of the network results in increasing the consumed energy since more communication tasks are performed inside the network. However, no significant change with the consumed energy occurred when EEORA is applied for both WSNs and DWSNs; so it is considered energy efficient routing algorithm. The third metric to evaluate the scalability is the energy-delay product that is attained by multiplying the delay with the amount of consumed energy within the network. It can be concluded that EEORA is effective to be applied within both WSNs and DWSNs since the network performance is not decrease while the network is being changed.

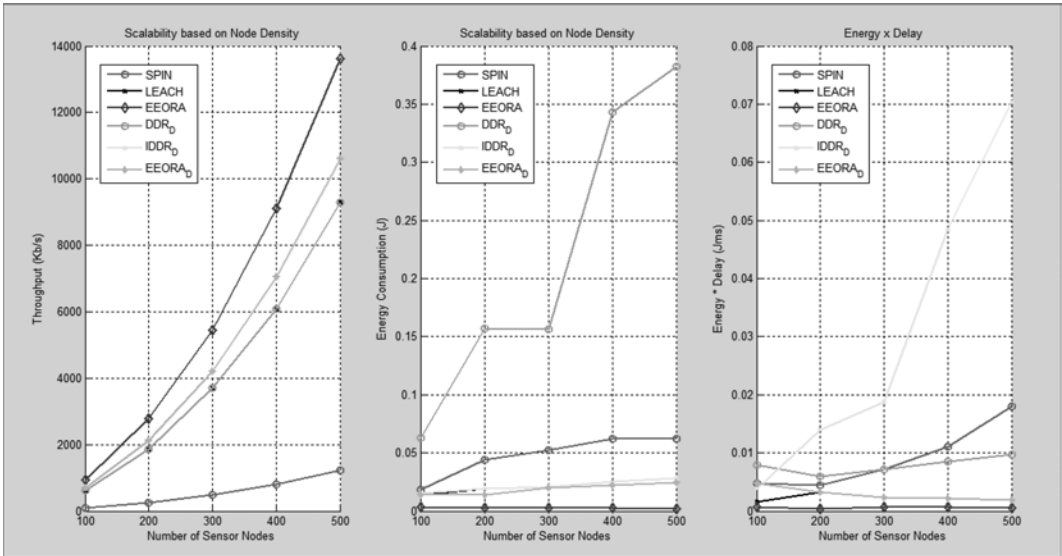


Figure 9. Scalability based on network density for both WSNs and DWSNs.

4 CONCLUSION AND FUTURE WORKS

As a conclusion; this paper introduced a novel routing algorithm that can be used in both WSNs and DWSNs. The proposed routing algorithm hybridizes the features of both flat and hierarchal routing techniques. Two phases were included in our proposed routing technique, which are, cluster formation and data communication phases. The clusters were formed based on improved IDDR algorithm and the data communication process is performed using new updated version of SPIN routing protocol. The proposed routing protocol, EEORA, was implemented in both WSNs and DWSNs in order to evaluate its performance in terms of different metrics, including; end-to-end delay, throughput, network life time and energy consumption. The results confirmed the effectiveness of EEORA to be applied in both WSNs and DWSNs. An improved performance in terms of all considered performance criteria was achieved for EEORA over SPIN, LEACH, DDR and IDDR protocols; this in turn gives EEORA more scalability to be applied as a routing technique in both WSNs and DWSNs.

REFERENCES

- Ahmad, A, K. Latif, N. Javaid, Z. A. Khan and U. Qasim. (2013). Density controlled divide-and-rule scheme for energy efficient routing in Wireless Sensor Networks. IEEE, 2013 26th Annual IEEE Canadian Conference on Electrical and Computer Engineering (CCECE).
- Al-Fares, M, Z. Sun and H. Cruickshank. (2009). A Reliable Multi-hop Hierarchical Routing Protocol in Wireless Sensor Network (WSN). 2009 Sixth International Conference on Information Technology: New Generations. 1604–1605.
- Ali and Parmanand (2015). Energy Efficiency in routing protocol and data collection approaches for WSN: A Survey. International Conference on Computing, Communication and Automation (ICCCA2015). 540–545.
- Gaoy, P and Y. Piao. (2014). DRRP: A Dynamically Reconfigurable Routing Protocol for WSN. IEEE. 460–465.
- Geetu and S. Juneja. (2012). Performance Analysis of SPIN and LEACH Routing Protocol in WSN. International Journal of Computational Engineering Research (ijceronline.com). 2 (5), 1179–1185.
- Guo, L and Q. Tang. (2010). An Improved Routing Protocol in WSN with Hybrid Genetic Algorithm. 2010 Second International Conference on Networks Security, Wireless Communications and Trusted Computing. 289–292.
- Khushboo, K and A.K. Daniel. (2015). Section Based Hybrid Routing Protocol for WSN using Artificial Bee Colony. 2015 International Conference on Advances in Computer Engineering and Applications (ICACEA). 887–892.
- Kodali, R and N. Sarma. (2013). Energy Efficient Routing Protocols for WSN's. 2013 International Conference on Computer Communication and Informatics (ICCCI-2013).
- Krishnaveni, R and J. Sutha. (2012). Analysis of routing protocols for wireless network. International Journal of Emerging Technology and Advanced Engineering. 2(11), 401–407.
- Li, X, W. Gang, L. Zongqi, Z. Yanyan. (2013). An energy-efficient routing protocol based on particle swarm clustering algorithm and inter-cluster routing algorithm for WSN. IEEE. 4029–4033.
- Malathi, L, M.K. Chandrasekaran and R.K. Gnana-murthy. (2012). A Novel Routing Protocol with Lifetime Maximizing Clustering Algorithm For WSN. IEEE. 925–930.
- Saleem, F, Y. Moeen, M. Behzad, M. A. Hasnat, Z. A. Khan, U. Qasim, N. Javaid. (2014). IDDR: Improved Density Controlled Divide-and-Rule Scheme for Energy Efficient Routing in Wireless Sensor Networks. Procedia Computer Science 34, 212–219.
- Singh, S, M.P Singh and D.K. Singh. (2010). Routing Protocols in Wireless Sensor Networks—A Survey. International Journal of Computer Science & Engineering Survey (IJCSSES. 1(2), 63–83.
- Srinath, R, A. Reddy and R. Srinivasan. (2008). AC: Cluster Based Secure Routing Protocol for WSN. Third International Conference on Networking and Services (ICNS'07).
- Ya, L, W. Pengjun, L. Rong, Y. Huazhong and L. Wei. (2014). Reliable Energy-Aware Routing Protocol for Heterogeneous WSN Based on Beaconing. ICACT2014. 109–112.
- Zhao, H, W. Zhou and Y. Gao. (2012). Energy Efficient and Cluster Based Routing Protocol for WSN. 2012 Eighth International Conference on Computational Intelligence and Security. 107–111.
- Zhixiang, D and Q. Bensheng. (2009). Three-layered Routing Protocol for WSN Based on LEACH Algorithm.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

MERS-CoV Disease Estimation (MDE) A study to estimate a MERS-CoV by classification algorithms

Manal Abdullah, M.S. Altheyab, A.M.A. Lattas & W.F. Algashmari

King Abdul-Aziz University, Jeddah, Saudi Arabia

ABSTRACT: Saudi Ministry of health has rolled a health regulation to people who arrive to perform religion rituals of Umrah and pilgrimage (Hajj). Seasonal loads enforce huge work in limiting epidemic spread especially with new viruses such as MERS-CoV that is extended in the region. Early disease diagnosis applications developed using techniques such as data mining. Data mining is a scientific field that interest in discovering patterns and relationships between data items. Algorithms that used in this field helps to increase decision effectiveness. Classification used widely in disease diagnosis, where patients of heart disease diagnosed correctly using classification algorithms that proven to be efficient in later applications. In this paper three classification techniques employed to diagnose patients with MERS-CoV through early syndromes. Naïve Bayesian, random forest, and support vector machine used to perform classification to patients' databases. Successively results were very good in predicting patients with coronavirus infection using random forest classifier. Additionally results were equivalent to medical statistics that exhibit elders as the most probable infected sector of people.

1 INTRODUCTION

The huge amount of data available in health sector and the need of extracting knowledge of these huge data make data mining techniques the most effective solution to process such amount of data and to extract knowledge. Data mining is the process of analyzing data and summarizing it into useful information, which can be used to increase revenue, cuts costs, or both (Bora, S.P. 2011). It is the process of finding correlations or patterns among dozens of fields in large relational data. Data mining consist of major elements: Extract information, store and manage data, provide access, analyze data and present data in a useful format.

Intersection between data mining with other fields such as machine learning, statistics, databases, pattern recognition, and knowledge acquisition for expert systems apply an ability of using different analytical tools and other techniques from intersectional fields (Mitra, S.); (Pal, S. K.); (Mitra, P. 2002). Analytical tools in data mining is used to extract relationships between data and their categories.

The advantage of intersection gives DM the power to be the most effective tool used in areas such as wireless sensor network (Zhu, H.); (Zhong, X.); (Yu, Q.); (Wan, Y. 2013), education (Kartiwi, M.); (Arundina, T.); (Omar, M. A.); (Gunawan, T. S. 2014), and business management (Romero, C.); (Ventura, S. 2010).

Healthcare uses data mining techniques for analyzing huge and complex volumes of data and extracting useful information. Activists generate a huge and complex volumes of data which make the analysis of it impractical. Data mining techniques generate an effective influence of such data (Soni, J.); (Ansari, U.); (Sharma, D.); (Soni, S. 2011). Data mining used in healthcare to provide many applications such as now a day, there are a lot of systems which help with basic operations. Clinical Decision Support System (Liu, X.); (Lu, R.); (Ma, J.); (Chen, L.); (Qin, B. 2015), health insurance fraud detection and identifying high-risk patient (Peng, Y.); (Kou, G.); (Sabatka, A.); (Chen, Z.); (Khazanchi, D.); (Shi, Y. 2006). A major applications of health care is disease detection that uses different algorithms to maintain early diagnosis, stopping infection spread and avoiding deaths if possible in addition to cost saving (Porter, T.); (Green, B. 2009). Widely used data mining technique is classification, this technique depends on assigning each record to a pre-defined category. Predicting a class of new entries used through predictive data mining techniques in prediction stage. Medically verity of classification algorithms utilized like: decision tree (Donghui Shi); (Jian Guan); (Zurada, J. 2015) and Fuzzy classifier (Shakir, M.); (Malik, A.S.); (Kamel, N.); (Qidwai, U. 2014).

In this paper, three classification techniques are used to diagnose patients with coronavirus or as

known MERS-CoV. Patients' dataset classified using naïve Bayesian, random forest, and SVM. The rest of paper is as follows: section two discusses in details classification techniques and the MERS-CoV, while in section three, the model that is used for classification is described. Section four is the experiment procedures and a review of results. Conclusions with future work of the paper is the last section.

2 LITERATURE REVIEW

For more than two decades, technology has been very important factor in medical areas it could help in preventing diseases by protecting against, decreases their risk and impact, reduce sickness duration, and improve the quality of care. In this section a brief about related studies in different fields.

2.1 *MERS-CoV*

Now a day in Saudi Arabia, most of government agencies and research centers focus on coronavirus (Middle East respiratory syndrome coronavirus (MERS-CoV)) such as Ministry of Health, Ministry of Agriculture and King Fahad Research Center. Family of coronaviruses can cause a wide range of diseases that range from common cold to Sever Acute Respiratory Syndromes (SARS). MERS-CoV is spreading around areas in all around the world and that imply huge interest around medical research community aiming to understand this newly known virus, associated symptoms, and treatment for it. Although transmission from human-to-human is not easy against transmission from animal-to-human, cases that is infected and not reported may spread this virus to people with close contact. The challenge with MERS-CoV is the big variation of symptoms that is mostly close to common cold symptoms with many other variation of diseases that may occur in cases and may not in other, some of patients have some unique symptoms while the other have no symptoms at all. A sample of 82 people that have the infection taken under research to figuring out the early syndromes related to this virus that shows the most important syndromes that is taken as the attributes of the classifier (Assiri, A.); (McGeer, A.); (Perl, T. M. 2014), (Assiri, A.); (Al-Tawfiq, J. A.); (Al-Rabiah, F. A. 2014), (Memish ZA); (Cotten M); (Watson SJ 2014). Another 10 cases added to cover the cases that can be found with no signs or symptoms of infection, taken under consideration the knowledge that was provided to us through doctors and World Health Organization (WHO) website. According to this, authors are working

to deploy data mining techniques in MERS-CoV diagnosis.

2.2 *Classification algorithms*

In this section, a review of some classification algorithms are presented. The algorithms adopted by this research are: Naive Bayes, Random forest, and Support Vector Machine SVM. Naïve Bayesian classifier works by limiting the input to categorical data and apply it to classification. A probability condition that formed after the observation of frequent data in the training dataset operated to guide the classification (Akkaya, G. C.); (Uzar, C. 2011). Many medical applications developed using naïve Bayesian such as: clinical decision support system CDSS that implemented to validate the state of patient at the moment and helping with early detection of any disease by examining patient symptoms with the data that is related to some disease and a recommendation of ways of prevent can be decided (Rane, A. L. 2015). An unexpected death in children with cardiac disease studied to be decreased where naïve Bayesian used to classify LQTS with known risk measurement (Qu, L.); (Vetter, V.L.); (Bird, G.L.); (Qiu, H.); (White, P.S. 2010).

Boosting and bagging are well known techniques of classification trees. In boosting, when early predictors make incorrect prediction, the successive trees give extra weight to points that incorrectly predicted. In bagging, each tree is constructed independently and using a bootstrap sample of the data set. Random forest changes the classification or regression trees constructed way. It added an extra layer of randomness to bagging and building each tree using different bootstrap sample of the data. Random forest splits in each node by the best divider among a subset of predictors. This approach performs an excellent work comparing with other classifier. It has only two parameters: the number of trees in the forest and the number of variables in the random subset at each node (Liaw, A.); (Wiener, M. 2002). Receiver Operating Characteristics (ROC) curve is utilized in evaluation by many researchers. Building model that predicts post-surgical death and kidney injury for patients after surgery (Turner, Z.); (Carroll, T.); (Brown, D. E. 2014). In detecting adverse Drug Events (ADEs) in Electronic Health Records (EHRs) (Zhao, Jing); (Henriksson, Aron); (Bostrom, Henrik 2015), the researchers apply a cascade of random forests to determine either the family of ADEs or the specific ADE. They conclude that proposed cascading scheme is more effective than a one-step prediction.

Support Vector Machine (SVM) deployed to classify binary variables with the availability

of using it in multiclass problems. It works by enlarging the gap between classes. Classification or regression make use of SVM by its ability to construct a hyper plane or multiple hyper planes in high dimensional space. This separates the data into parts of the original input space (Tomar, D.); (Agarwal, S. 2013). Healthcare applications used this technique in many areas to achieve different goals such as defining a mortality prediction model that help in making a suitable decision of a patient. It needs and demands a fast and extensive care (Liu, J.M.); (You, M.); (Li, G.Z. 2013). Another application in healthcare that provide a heart failure early diagnosis in addition to risk analysis made by using SVM classifier incorporated with other classifiers such as naïve Bayesian (Fu, X.); (Ren, Y.); (Yang, G. 2011).

3 RELATED WORK

Utilization of data mining techniques increased over the years to cover group of medical applications in order to enhance services quality in this field. A brief overview of these applications introduced in this section. (Nishara Banu, M.A.); (Gomathy, B. 2014) from Bannari Amman Institute of Technology published a paper about a heart disease prediction system using clustering and classification algorithms. They use k-means to cluster relevant data and MAFLA to mining maximal frequent patterns in the database then apply classification with ID3 and C4.5, this provides the 89% of accuracy. (Jabbar, M.A.); (Deekshatulu, B.L.); (Chandra, P. 2015), from Telangana in India, proposed a method to enhance Naïve Bayes algorithm by using discretization and feature subset selection measure like chi-square, gain ratio. One-R and genetic search proposed to reduce the number of diagnostic test to be taken individually by distortive data.

Another research contains an improved study of Heart Disease Prediction System using three classification techniques: naïve bayesian, decision tree, neural network by (Dangare, C. S.); (Apte, S.S. 2012) from walchand Institute of Technology. (Bruser et al 2013) suggested another study about possibility of automatic detection of atrial fibrillation from cardiac vibration signals that recorded by unobtrusive bed mounted sensors. By applying seven machine learning techniques and based on the given set of time-frequency-domain and time-domain features, the random forest classifier gives best result. (Jyoti Soni); (Ujma Ansari 2011) from Raipur Institute of Technology write a paper about predictive data mining for medical diagnosis. They focus in using different data mining algorithms for an intelligent and effective prediction of heart disease.

4 SYSTEM MODEL

Diagnosing patients with MERS-CoV through early syndromes is mainly the concern of this paper, where this gives an ability to limit infection spreading especially through seasons such as Umrah and Hajj. Helping new doctors in diagnosis aimed through this research, where accurate diagnosis is necessary. Suggested system flow chart is shown in ‘Figure 1’.

The stages shown represents five major steps of the system. First is the dataset collected from UCI and MERS-CoV cases recorded from number of medical analytical papers that is concerned with early symptoms of that disease. Issues such as irrelevant data, redundant information, noisy and unreliable data that can increase the difficulty of

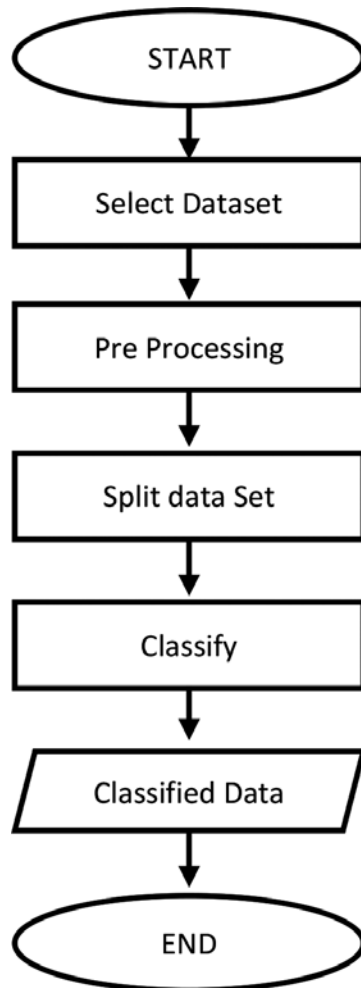


Figure 1. System model.

prediction are needed to be handled using some preprocessing techniques. At preprocessing stage, the applied dataset will be prepared through filtering steps: removing irrelevant attributes, replacing missing values, transformation to ARFF file format to be suitable to apply by WEKA software package (Sharma, T.C.); (Jain, M. 2013). Then, data divided into training and testing sets by different splitting percentages. Both training and testing sets manipulated with many classifier algorithms. This is the main stage in the model when the classifier algorithm applying to the data set to estimate the probability of the disease infection by interning a record and provide the classification for it. Furthermore, in testing classification algorithm try to achieve or predict an identical class to pre-classified row.

5 EXPEREMENTAL SETUP AND RESULTS

At the beginning of this section a description of the dataset will be presented. A dataset used in this research contains 322 records. The set has 92 infected cases while 230 uninfected cases. Each record contains 24 attribute as shown in Table 1 to help increase the accuracy of prediction. The algorithm will classify the record and assign it to one of two categories: Low and High according to the probability of its features or based on the classification rules.

Three supervised learning algorithms implemented in this research, naïve Bayesian and random forest, SVM. These classifier algorithms are used for evaluating the performance exploited to choose the best classifier by three evaluation measurement parameters. TP rate that present a positive set that is classified correctly (Sharma, T. C.); (Jain, M. 2013), in addition to FP rate that shows incorrectly classified negative tuples (Sharma, T. C.); (Jain, M. 2013). Receiver Operating Characteristic (ROC) used also as another parameter that used as a gaudiness in training by error measurement (Sharma, T. C.); (Jain, M. 2013). Researchers used two test methods that are the percentage split point and the 10-fold cross validation procedures. Many splitting ratios that ranges from 30% to 70% are used and illustrated in Table 2.

Best classifier chosen among others based on ROC since results were close as shown in 'Figure 2'. 'Figure 2' and 'Figure 3' illustrated that random forest perform an excellent result against the other classifiers within 70% splitting point. Naïve Bayesian follows to perform a good prediction capability. Minimum correctly classified instances are given by naïve Bayesian classifier at 60% splitting point. It is also noticeable that minimum TP ratio is found at random forest with 30% splitting point while the maximum value presented at naïve Bayesian with 10-foldcross validation.

Table 1. Attributes description.

Attribute	Type	Range
Fever	Nominal	High—Mid—Low
Age	Numeric	3-96
Sex	Numeric	0-1
Fasting blood sugar	Numeric	0-1
Heart disease	Nominal	Present—Absent
Chronic kidney	Nominal	CKD—not CKD
Chills	Nominal	Yes - /No
Dry	Nominal	Yes—No
Productive	Nominal	Yes—No
Shortness of breath (SOB)	Nominal	Yes—No
Sore throat	Nominal	Yes—No
Runny nose	Nominal	Yes—No
Abnormal pain	Nominal	Yes—No
Nausea	Nominal	Yes—No
Vomiting	Nominal	Yes—No
Diarrhea	Nominal	Yes—No
Myalgia	Nominal	Yes—No
Headache	Nominal	Yes—No
Hypertension	Nominal	Yes—No
Chronic lung	Nominal	Yes—No
Obesity	Nominal	Yes—No
Smoking	Nominal	Yes—No
Chest pain	Nominal	Yes—No
MERS-CoV	Nominal	Yes—No

Table 2. Results of classification techniques.

	NB	RF	SVM	Experiment choice
TP ratio	0.889	0.862	0.729	30% splitting point
FP ratio	0.246	0.33	0.649	
ROC	0.885	0.915	0.54	
TP ratio	0.896	0.886	0.793	40% splitting point
FP ratio	0.19	0.247	0.507	
ROC	0.885	0.907	0.643	
TP ratio	0.882	0.882	0.789	50% splitting point
FP ratio	0.194	0.266	0.461	
ROC	0.883	0.91	0.664	
TP ratio	0.876	0.876	0.783	60% splitting point
FP ratio	0.205	0.266	0.413	
ROC	0.872	0.906	0.685	
TP ratio	0.897	0.887	0.773	70% splitting point
FP ratio	0.163	0.246	0.413	
ROC	0.907	0.942	0.68	
TP ratio	0.907	0.898	0.832	10 cross validations
FP ratio	0.157	0.217	0.357	
ROC	0.905	0.936	0.738	

An important point in this research is to determine threshold level of age as an attribute with high impact on classification process. Age classified into two categories: old and young, an enhancement of category range performed consequently and classifiers implemented with each

range. Dataset updated by changing age attribute first into nominal with two predefined values. After adopting new value, dataset exposed to already chosen classifiers: naïve Bayesian, random forest, and SVM Results show that elderly have a high probability of infection with coronavirus than younger. Table 3 shows different classification of the age range for old people with percentage of correctly classified instances using different classifiers that is illustrated in ‘Figure 4’. ‘Figure 5’ represents threshold curve of different old age range using NB classifier with 70% splitting point.

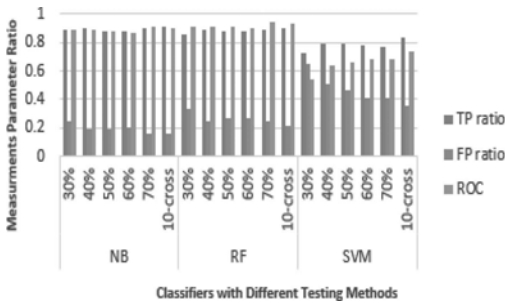


Figure 2. Classifiers results.

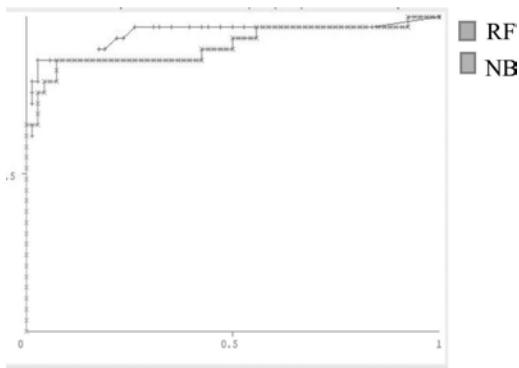


Figure 3. Classification measurement using ROC.

Table 3. Results of changing threshold level.

Age-range	Splitting point	NB	RF	SVM
26-	30%	88.89%	87.56%	88.67%
	70%	89.69%	89.69%	88.66%
31-	30%	88.44%	86.22%	87.11%
	70%	89.69%	87.63%	88.66%
50-	30%	88%	87%	87.11%
	70%	90%	88%	88.66%
70-	30%	88.89%	87.11%	87.56%
	70%	90.72%	89.69%	88.66%

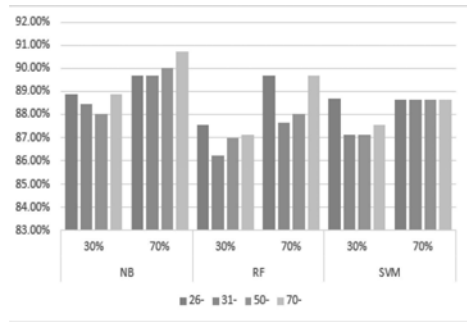


Figure 4. Threshold of age.

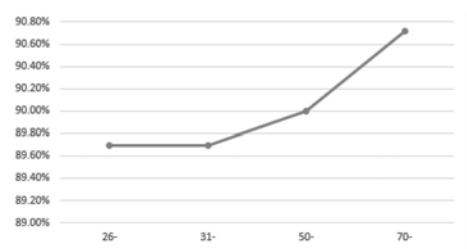


Figure 5. NB Result with different age ranges.

6 CONCLUSION AND FUTURE WORK

Disease estimation represents an important and considerable part of medical data mining studies, where many researchers exploit techniques such as classification to enhance and conduct their theory. In this work, applying a dataset to a group of classifiers: naïve Bayesian Classifier, random forest classifier, and SVM help in estimating the MERS-CoV disease. Measuring the performance of each algorithm are conducted using ROC where the Random Forest shows the best results. By changing the splitting percentage to 70%, Random Forest Classifier presents good performance with ROC measurement to 0.942%. Another important conclusion that is tested by this research that mostly elders between 50 to 96 years old possibly exposed more than young people to coronavirus infection. While the proposed model extracts an amazing result, it can be enhanced by applying larger data set as future of this work that taken under consideration larger set of patients and cases such as pregnant women with implying different classification algorithms.

REFERENCES

Akkaya, G.C., & Uzar, C. (2011). Data Mining: Concept, Techniques And Applications. GSTF Business Review (GBR), 1(2), 47.

- Assiri, A., Al-Tawfiq, J.A., Al-Rabeeah, A.A., Al-Rabiah, F.A., Al Hajjar, S., Al-Barrak, A.,... & Makhdoom, H.Q. (2013). Epidemiological, demographic, and clinical characteristics of 47 cases of Middle East respiratory syndrome coronavirus disease from Saudi Arabia: a descriptive study. *The Lancet infectious diseases*, 13(9), 752–761
- Assiri, A., McGeer, A., Perl, T.M., Price, C.S., Al Rabeeah, A.A., Cummings, D.A.,... & Madani, H. (2013). Hospital outbreak of Middle East respiratory syndrome coronavirus. *New England Journal of Medicine*, 369(5), 407–416.
- Bora, S.P., “Data mining and ware housing,” in *Electronics Computer Technology (ICECT)*, 2011 3rd International Conference on, vol.1, no., pp.15, 8–10 April 2011.
- Bruser, C., Diesel, J., Zink, M.D., Winter, S., Schauerte, P., & Leonhardt, S. (2013). Automatic detection of atrial fibrillation in cardiac vibration signals. *Biomedical and Health Informatics, IEEE Journal of*, 17(1), 62–171.
- Dangare, C.S., & Apte, S.S. (2012). Improved study of heart disease prediction system using data mining classification techniques. *International Journal of Computer Applications*, 47(10), 44–48.
- Donghui Shi; Jian Guan; Zurada, J., “Cost-Sensitive Learning for Imbalanced Bad Debt Datasets in Healthcare Industry,” in *Computer Aided System Engineering (APCASE)*, 2015 Asia-Pacific Conference on, vol., no., pp.30–35, 14–16 July 2015
- Fu, X., Ren, Y., Yang, G., Pan, Q., Gong, S., Li, L.,... & Ning, G. (2011, September). A computational model for heart failure stratification. In *Computing in Cardiology*, 2011 (pp. 385–388). IEEE.
- Jabbar, M.A.; Deekshatulu, B.L.; Chandra, P., “Computational intelligence technique for early diagnosis of heart disease,” in *Engineering and Technology (ICETECH)*, 2015 IEEE International Conference on, vol., no., pp.1–6, 20–20 March 2015
- Jyoti Soni, Ujma Ansari, Dipesh Sharma and Sunita Soni. Article: Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction. *International Journal of Computer Applications* 17(8):43–48, March 2011.
- Kartiwi, M., Arundina, T., Omar, M.A., & Gunawan, T.S. (2014, November). S-Rater: Data mining application in Islamic financial sector. In *Information and Communication Technology for The Muslim World (ICT4M)*, 2014 The 5th International Conference on (pp. 1–5). IEEE.
- Liaw, A., & Wiener, M. (2002). Classification and regression by random Forest. *R news*, 2(3), 18–22.
- Liu, J.M., You, M., Li, G.Z., Wang, Z., Xu, X., Qiu, Z.,... & Chen, S. (2013, July). Cough signal recognition with Gammatone Cepstral Coefficients. In *Signal and Information Processing (ChinaSIP)*, 2013 IEEE China Summit & International Conference on (pp. 160–164). IEEE.
- Liu, X.; Lu, R.; Ma, J.; Chen, L.; Qin, B., “Privacy-Preserving Patient Centric Clinical Decision Support System on Naive Bayesian Classification,” in *Biomedical and Health Informatics, IEEE Journal of*, vol. PP, no.99, pp.1–1.
- Memish ZA, Cotten M, Watson SJ, et al. Community case clusters of Middle East respiratory syndrome coronavirus in Hafr Al-Batin, Kingdom of Saudi Arabia: a descriptive genomic study. *Int J Infect Dis* 2014; 23:63–8.
- Mitra, S., Pal, S.K., & Mitra, P. (2002). Data mining in soft computing framework: a survey. *IEEE transactions on neural networks*, 13(1), 3–14.
- Nishara Banu, M.A.; Gomathy, B., “Disease Forecasting System Using Data Mining Methods,” in *Intelligent Computing Applications (ICICA)*, 2014 International Conference on, vol., no., pp.130–133, 6–7 March 2014.
- Peng, Y., Kou, G., Sabatka, A., Chen, Z., Khazanchi, D., & Shi, Y. (2006, October). Application of clustering methods to health insurance fraud detection. In *Service Systems and Service Management*, 2006 International Conference on (Vol. 1, pp. 116–120). IEEE.
- Porter, T., & Green, B. (2009). Identifying diabetic patients: a data mining approach. *AMCIS 2009 Proceedings*, 500.
- Qu, L., Vetter, V.L., Bird, G.L., Qiu, H., & White, P.S. (2010, December). A Naïve Bayes classifier for differential diagnosis of Long QT Syndrome in children. In *Bioinformatics and Biomedicine (BIBM)*, 2010 IEEE International Conference on (pp. 433–437). IEEE.
- Rane, A.L. (2015, January). Clinical decision support model for prevailing diseases to improve human life survivability. In *Pervasive Computing (ICPC)*, 2015 International Conference on (pp. 1–5). IEEE
- Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 40(6), 601–618.
- Shakir, M.; Malik, A.S.; Kamel, N.; Qidwai, U., “Intelligent Fuzzy Classifier for pre-seizure detection from real epileptic data,” in *Science and Information Conference (SAI)*, 2014, vol., no., pp.276–279, 27–29 Aug. 2014
- Sharma, T.C., & Jain, M. (2013). WEKA approach for comparative study of classification algorithm. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(4), 1925–1931.
- Soni, J., Ansari, U., Sharma, D., & Soni, S. (2011). Predictive data mining for medical diagnosis: An overview of heart disease prediction. *International Journal of Computer Applications*, 17(8), 43–48.
- Terner, Z., Carroll, T., & Brown, D.E. (2014, October). Time series forecasts and volatility measures as predictors of post-surgical death and kidney injury. In *Healthcare Innovation Conference (HIC)*, 2014 IEEE (pp. 319–322). IEEE.
- Tomar, D., & Agarwal, S. (2013). A survey on Data Mining approaches for Healthcare. *International Journal of Bio-Science and Bio Technology*, 5(5), 241–266.
- Zhao, Jing; Henriksson, Aron; Bostrom, Henrik, “Cascading adverse drug event detection in electronic health records,” in *Data Science and Advanced Analytics (DSAA)*, 2015. 36678 2015. IEEE International Conference on, vol., no., pp.1–8, 19–21 Oct. 2015
- Zhu, H., Zhong, X., Yu, Q., & Wan, Y. (2013, January). A localization algorithm for mobile wireless sensor networks. In *Intelligent System Design and Engineering Applications (ISDEA)*, 2013 Third International Conference on (pp. 81–85). IEEE.

E-learning standards

Manal Abdullah

King Abdulaziz University, KSA

Nashwa Abdel Aziz Ali

College of Computing and Information Technology, Information System Department,
Arab Academy for Science, Technology and Maritime Transport (AASTMT), Egypt

ABSTRACT: Standardization is the process of developing “specifications.” Specifications can be identified as characteristics and attributes of their development, prior to getting an approval from authorized organizations. At this concept, this paper presents the issue of E-learning standards and specifications that are developed by international E-learning organizations over latest years. The paper discusses all aspects of E-learning standards and specifications throughout approaches, dimensions, systems, and tools which are gaining special attention in the last years. The main contribution in this paper is identifying E-learning standardization, shedding the light to E-learning standards; providing a summary of e-learning standards and specifications that are reaching a state of applicability and certified in the general committees.

Keywords: E-learning; standrization; Standards; Sersonalization; Adaptaitaon; E-learning quality

1 INTRODUCTION

The process of standardization is referred to be experimental, incomplete and more rapidly evolving. Standards, on the other hand, are much more conclusive, complete, and evolve slowly. Defining E-learning standards is used consistently as rules, guidelines, or definitions of characteristics, to ensure that materials, products, processes and services are fitted for their purpose. In the context of e-learning technology, standards are generally developed to be used in systems, design and implementation for the purposes of ensuring interoperability and reusability. These attributes should apply to both the systems themselves, content and meta-data they managed. In addition, these standards should capture general acceptance, can serve regulatory purposes, and be used to achieve E-learning efficiency and quality outcomes. One of the main benefits of standardized E-learning is to establish the functionality of interoperability via multi systems and re-use of learning objects (interoperability of resources). Interoperability and developing efficiency can be called as the two main purposes of standardization of E-learning [1]. Standardization differentiates between common types of standards and specifications as follows:

Formal standards: It is also known as “de-jure standards”. The International Organization for Standardization (ISO) and International

Electrotechnical Commission (IEC) are developed in agreement processes by the official standardization organizations.

Community specifications: They are developed by communities or forums. They are open specifications which available in public. The Institute of Electrical and Electronics Engineers, Inc. (IEEE) and the World Wide Web Consortium (W3C) are E-learning examples communities.

Industrial specifications: They are developed by closed or open specifications and are available for branches of industrial consortia.

Organizational specifications are internally closed specification developed by organizational industrial or community specifications with globally agreement. Microsoft Windows is an example of proprietary organizational, industrial or community standards.

E-learning standards have different levels of users and providers to address and support in their needs, interests and preferences.

The major advantages of developing E-Learning standards are [2]:

- Durability—no need to modifying standard when issue new system version.
- Interoperability—operating across vary hardware, operating systems, web browsers and multi learning management systems.
- Accessibility—indexing and monitoring.

- Reusability—enabling use by different development tools.

E-learning standards can be viewed from two main viewpoints. First is to use standards as a mean for creating adaptive learning scenarios. These can be imported and run in the LMS to deliver learners with individualized learning experiences. While the second view supports standards that concern with information on the experiences and state of learning of each learner. This vision is supported to use SCORM and xAPI as a provider of information on the learning progress of each learner [2].

The rest of this paper is organized as: section 2 represents the categories of E-learning standards. Section 3 describes the various learning styles. In section 4, the research gives an overview of adaptation e-learning. At next section the research browses the highly cited quality standards then browses the organizations that issue e-learning standards and finally it presents the trends of e-learning.

2 CATEGORIES OF E-LEARNING STANDARDS

Categories of E-learning standards may support the multidimensional concept such as learning content and learning objects, processes and the business units of organization.

The most important purpose of learning standards can be divided into five categories as following [4]:

1. **Metadata:** Learning content and catalogs must be labeled in a consistent way to support the indexing, storage, discovery (search), and retrieval of learning objects by multiple tools across multiple repositories. Several initiatives are creating metadata standards: The Learning Object Metadata, or LOM of IEEE Learning Technology Standards, and the Dublin Core Metadata.
2. **Content Packaging:** Content packaging specifications and standards allow courses to be transported from one learning system to another. The initiatives dealing with content packaging include: The IMS Content Packaging specification [8], the IMS Simple Sequencing specification, the ADL Sharable Content Object Reference Model (SCORM).
3. **Learner Profile:** Learner profile information can include personal data, learning plans, learning history, accessibility requirements, certifications and degrees, assessments of knowledge and the status of participation in current learning. The most important effort to standardize learner profile information is the IMS Learner Information Package (LIP) specification.

4. **Learner Registration:** Learner registration information allows learning delivery and administration components to know what offerings should be made available to a learner, and provides information about learning participants to the delivery environment. There are two initiatives currently dealing with these requirements: The IMS Enterprise Specification, and the Schools Interoperability Framework which supports the exchange of this type of data in K-9 environment.
5. **Content Communication:** When content is launched, there is a need to communicate learner data and previous activity information to the content. Work going on is the ADL's Sharable Content Object Reference Model (SCORM) project based on the CMI specification of the Aviation Industry CBT Committee.

Categorizing E-learning standards can be identified by three main dimensions of standards: Types, domains, and entities of e-learning standards. In the following subsections a brief of the three dimensions is introduced [3].

2.1 Types of e-learning standards

Three types of E-learning standards can be identified. These attribute the main functions of E-learning standardization:

1. **Implementation Standards:** These types support the functionality of interoperability within all E-learning domains. Metadata, Architecture, Infrastructure, Interface standards are example of implementation standards.
2. **Conceptual Standards:** These types of standards support the functionality of quality development by providing reference models.
3. **Level Standards:** These types focus on identifying and addressing the quality level of E-learning systems. So, they usually used for certifications purposes. Figure 1 shows the relationship between purposes of E-learning standards and its types.

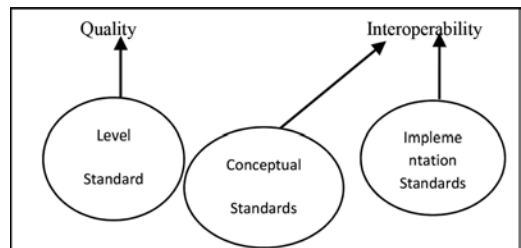


Figure 1. The relationship between the purpose of e-learning standards and its types.

2.2 Domains of E-learning standards

E-Learning standards may cover one or more domain of the following e-Learning domains:

1. Meaning: This domain focuses on the general concepts of understanding and deals with: disciplines semiotics, pragmatics, semantics, etc...
2. Quality: This domain covers all aspects of the quality management including developing, assurance and deals with results, processes, and potentials.
3. Didactics: This domain focuses on issues concerning with pedagogical questions: methods, learners, and learning environments.
4. Learning technology: This domain focuses on all technological solution that developed for learning objectives and purposes and deals with: data exchange, interfaces, and accessibility questions.
5. Learning content: This domain covers all aspects for e-Learning objects and deals with: resources, aggregation, and packaging.
6. Context: The main purpose of this domain is to include all the disciplines and information regarding to e-Learning and to deal with: rights and laws.

2.3 Entities of E-learning standard

E-Learning standards may correspond to more than one entity in combination. Six entities can be addressed as the following:

1. Learning environment: This entity concerns with the entire organizational environment including structure, infrastructure, services and all processes.
2. Roles: it deals with learner, teacher, and tutor that represent the defined attributes within the E-Learning solution.
3. Methods: It concerns with the used methods defined for and used within an e-Learning environment.
4. Learning systems: It deals with all technological and conceptual issues including the architecture within the system.
5. Learning resources: It covers all the components of the learning system.
6. Practice: It deals with all information experiences and knowledge that recognized within the usage of an e-Learning offer.

3 STANDARDS FOR E-LEARNING STYLES

A learning style is referred to characteristics, strengths and preferences that represented the

learner's method when received information through learning process. This section gives a notice about most cited learning styles [3].

3.1 Kolb learning style indicator

It is an indicator based on "Experiential Learning Theory" which considers experience of a learner as an important factor in learning. Therefore, it discussed two kinds of experiences namely grasping and transforming experiences. Grasping consists of two sub categories namely concrete and abstract conceptualization. Similarly, the transforming experience has two sub categories termed reflective observation and active experimentation.

3.2 Fleming VAK model

The VAK learning styles model suggests that most people can be divided into one of three preferred styles of learning namely Visual, Auditory and Kinesthetic.

- Visual learning style involves the use of seen or observed things, including pictures, diagrams, demonstrations, displays, handouts, films, flip-chart, etc.
- Auditory learning style involves the transfer of information through listening: to the spoken word, of self or others, of sounds and noises.
- Kinesthetic learning involves physical experience—touching, feeling, holding, doing, and practical hands-on experiences.

3.3 Myers Briggs Type Indicator (MBTI)

The learning style assessment by MBTI is resolved using different aspects. The dimensions of this learning styles are as follows:

- Sensing Judging & Perceiving: Attention towards the external world/things or internal world/things.
- Thinking & Feeling: Perceive world directly or perceive through impressions/imaging possibilities.
- & Intuition: Learners taking decisions through logic or through mere human values.
- Extroversion & Introversion: Learner looking the world as a structured, planned environment or as a spontaneous environment.

3.4 Felder–Silverman index of learning styles

This learning style model is often used in technology enhanced learning and designed for traditional learning. Additionally, it defines the learning style of a learner in more detail and distinguishing between preferences on four dimensions as:-

- Active/Reflective: Active learners learn by doing something with information. Reflective learners learn by thinking about information.
- Sensing/Intuitive: Sensing learners favor to take in info that is concrete and practical. Intuitive learners favor to take in info that is original, abstract and oriented towards theory.
- Visual/Verbal: Visual learners prefer visual presentations of material—pictures, diagrams, flowchart and graphs. Verbal learners prefer explanations with words—includes both written and spoken.
- Sequential/Global: Sequential learners prefer to organize information in a linear, orderly fashion.

Global learners prefer to organize information more holistically and in a seemingly random manner without seeing connections.

Since most learners fall in the category of either active or reflective for the first dimension, this model is more suitable to evaluate the learners in an E-learning environment.

3.5 Grasha-Riechmann student learning style scales

It focuses on students' interactions amongst their peers. GRSLSS has a teaching style survey that instructors can complete to see how their instruction matches or not with learners so they can adapt and modify to meet more learner's needs. The learning styles scale consists of 6 primary learning styles are: Avoidant, Collaborative, Competitive, Dependent, Independent, and Participant. The survey itself consists of 60 items, with 10 questions each. They are averaged together to measure dominance in one or more of the six measured learning styles.

4 E-LEARNING PERSONALIZATION STANDARDS

Personalization is wider than just individualization or differentiation in what does learner prefer about what is learnt, when and how is learnt. Personalized Learning is the tailoring of learning environment to meet the needs and preference of individual learners [7].

4.1 Personalized e-learning Elements

A variety of elements are involved in personalized e-learning to customize learning process. Key elements that meet the learner goals, and personalized E-learning process include:

- The pace of learning, pedagogy, curriculum and instructional approach, activities that draw upon the student's preferences, skills and knowledge.

In essence, personalized e-learning environment enables learners to choose what they want to learn when they want, and even the learning method. This almost leads to improve learning outcomes.

4.2 Personalized E-learning goals

The main goal to achieve continuing support for every learner at each individual level and monitor his behavior throughout learning process is being material:

- Be modified to each learner,
- Be generative rather than pre-composed
- Be scalable to learning process levels without additional cost.

Personalization mechanism is based mainly on feedback. It can either in a form of explicit or implicit and can be in manually or automatically processed technique. It should be an integral part of the learning system.

4.3 Personalized E-learning methods

Personalized E-learning should be widely focus on enhancing learning and thinking interaction at learner behavioral and physiological levels. Knowledge driven model for personalization E-learning solutions is a part from sophisticated, stylish multimedia delivery of learning process.

Another centric aspect of enhancing E-learning is an adoption of what is the preferred pace and expertise of learner. The future direction of E-learning is shifting from a content-oriented approach to a knowledge synthesis approach.

Two main approaches are considered in personalized E-learning methods [7]:

Knowledge model and concept map approach provide successful personalization of E-learning process by designing a platform to interact a continuous dialogue between the learner and knowledge resources. Concept maps are used as the graphical representations of knowledge to draw both the learning concepts and its relationships with a human oriented approach.

Tacit knowledge is highly subjective in nature, as it is developed by learner and based on his cognitive and related to context of a specific situation. The Tacit (T)-model is used to capture the sequential set of steps that a Subject Matter Expert (SME) would handle to achieve a task or to make a decision. As an important conceptual structure of the T-model, it is identified by using a formal concept analysis (FCA) which is used to restructure and formalize the T- model.

Explicit knowledge is the knowledge that is objective in nature, easily expressed and shared. Explicit knowledge is modeled in the subject

domain master map (M-map) which can be hyper linked to each knowledge node in the T-map. The M-map is formulated on the concept of learning dependency, which is defined as a dynamic cognitive and pedagogical centered approach for the mapping of a course structure.

Personalization in E-learning system can be achieved through two levels of personalization. Level 1 allows the personalization of learning contents and structure of the course according to a given personalization strategy and level 2 defines the personalization strategy. Teacher has to choose and apply the personalization strategy which matches the learner's characteristics and specifics of the courses in two steps, first, teacher selects a subsets of personalization parameters for given courses and then, combines the selected

Table 1. Personalization parameters and its values.

Personalization parameter	Values
Information seeking tasks	Tracking activities
Knowledge level	Learner background, values are: beginner, intermediate, and advanced.
Learning goals	Knowledge, comprehension, application
Media preference	text/image, sound, video, and simulation
Language preference	Represent learning objects in the learner's preferred language
Kolb learning cycle	
Honey-Mumford learning	
Felder-Silverman learning	
Gransh learning style	
Participation	Too much, acceptable, not enough
Progress on task	Large, small
Feedback	Significant, medium, low
Motivation level	Components of motivation instructions (Attention, Relevance, Confidence, and Satisfaction). Set of values: low, moderate, and high
Navigation level	Varying between breadth -first and depth first.
Cognitive traits	Low, high working memory capacity & Low, high inductive reasoning ability & Low, high information processing speed & Low, high associative learning skill
Pedagogical approach	Objectivist, competency

personalization parameters and decides how the learning material can be composed with respect to each possible value of the personalization parameters. There are 16 personalization parameters of E-learning scenarios such as: Information seeking task, learner's level of knowledge, learning goals, language preference, etc. Table (1) shows personalization parameters and its values according to OSPS standard (Ontology for Selection of Personalization Strategy)

5 E-LEARNING ADAPTATION STANDARDS

The crucial criterion to providing a responsive learning environment that engages motivates and inspires learners, and through this leads to higher learner satisfaction, is Adaptation.

In the context of this paper, a learning environment is considered adaptive if it is capable of: monitoring the activities of its users; interpreting these on the basis of domain-specific models; inferring user requirements and preferences out of the interpreted activities, appropriately representing these in associated models; and, finally, acting upon the available knowledge on its users and the subject matter at hand, to dynamically facilitate the learning process [8, 9, 10]. There are numerous adaptive E-Learning standards that adopted analyzing various specifications such as IMS1, ADL, SCORM2 and AICC.

- Adaptation-oriented Domain Modelling

Current standards and concepts for educational metadata focus on content-centered approaches and models of instructional design. Standards focus on search, exchange and re-use of learning material, often called content items, learning objects or training components. Examples of ALEs that extend existing standards include OPAL, OLO and KOD, among others.

- Learner and Group Modelling

Learner modelling in existing standards addresses all related specifications to the learner's model, or profile, over time. An example of these type of specifications is the IMS Learner Information Package specification, which incorporates the results of "top-level" educational activities, and other static information about the user (e.g., demographic).

- Adaptation Modelling

Two complementary issues of modelling examine the behavior of any adaptive system: the specification of adaptation logic, and the specification of adaptation actions. The prior is responsible for relating information available in one or more models and assesses whether adaptation is required. The latter refers to specifying the actions needed

for given adaptation should be achieved. The two approaches include simple rule-based engines, case-based reasons, etc. An adaptation logic, adaptation actions constitute a well-researched, especially as far as Adaptive Hypermedia Learning Systems are concerned. Furthermore, recent research used an XML language to define and declare adaptation actions. Of the existing standards, the only one that supports the explicit representation of dynamic behavior on behalf of the system is the IMS Learning Design (LD) specification.

- Standardization at Adaptation Components and Services Level

This type of standards is concerned with utilizing adaptation-oriented components services to/from the “outside world” PLS and Knowledge Tree. Both of these adaptation techniques are constructed to existing source content and functionality (such as Learner Management, Collaborative Tools, and Testing Services). PLS integrates with learning management system ADL SCORM based and work within existing courseware. PLS is completely static and achieves adaptation via adaptation services.

Knowledge Tree framework, on the other hand, is designed to facilitate interoperation and reuse at the level of distributed, reusable learning activities. Furthermore, it works on run-time communication and interoperation standards, standardizes methods to support aspects of the adaptation learning process that can exchange information throughout LMS. Knowledge Tree allows for different kinds of portals—some can be as static as existing CMS, but the others can be adaptive.

6 E-LEARNING MANAGEMENT SYSTEMS (LMS) STANDARDS

E-learning Management Systems is a software that records, tracks and monitors all activities of learner. In other words, these systems are friendly structure and foundation for users of e-learning, handle learning and training process automatically. A powerful and comprehensive learning management system provides highly performance management of learning process. It helps learners assess their training and plan their next steps for learning [11].

All LMSs have their own specifications and properties, but represent it in different forms. Some of these systems enable users throughout learning process management tools. Some others provide learners to choose learning object based on their needs or choose form list of courses. In addition, these systems supply learners with educational activities in visual form.

LMSs almost provide the following functions:

- **Structure:** centralization and interoperability by enabling easy and efficient navigation via interfaces
- **Assessment:** creating and administrating assessments, storing assessment data, also includes all functions related to assessment results.
- **Tracking:** tracking learning functions into one system.
- **Security:** prevent unauthorized access to courses, learner account, and other administrative facilities.
- **Registration:** assigning to courses, activities by learners or instructors.
- **Delivery:** on-demand delivery of learning content and experiences to learners.
- **Interaction:** learner interact throughout all administrative tools as well as between communicative content and the LMS (i.e. SCORM content).
- **Reporting:** extract and present all information about learners and courses.
- **Record keeping:** store and maintain data about learners including their portfolio.
- **Reuse:** search and compose contents for delivery in different learning tracks.
- **Personalization:** match learner preferences with corresponding contents
- **Integration:** exchange learner and content data via multi systems (i.e. content management systems).
- **Administration:** centralized management of all involved functions.

The following subsection present the most common standards developed to achieve the previous functionalities.

6.1 SCORM standard

SCORM is a technical standard that was created and developed by ADL. This standard supports the following keys as high-level requirements:

Availability, adaptability, economic, durability, interoperability and reusability. In another words SCORM is a collection of related documents. Three main documents of SCORM [12]:

1. Content aggregation model
2. Runtime environment
3. Arrange and conduct

In fact, SCORM is a highly level set of fundamental characteristics and e-learning content standards, technologies and related services. SCORM 2004 introduced a complex idea called sequencing, which is a set of rules that constrain a learner to fix his paths and bookmark learning object. The standard uses XML to encode a file that describes the components and resources.

6.2 Metadata and Interoperability standards

For economic and efficient benefits, learning organization must be associated and supported by metadata and interoperability standards. Metadata is information about information that considered as identification or the main characteristics of these information [11]. By other words, Metadata is structured information that describes, explains, locates, and makes information easier to retrieve, use and manage. It includes information about subject, author, title, book title, author, publisher, edition and other necessary information in libraries are suitable examples of metadata. MIME is also a part of the standard metadata that is related to information posted on the internet and informs receiver about information and software required to process it.

Interoperability means that standards support different systems or different components or layers of systems. Metadata standards play a role of Interoperability standards.

Organizations such as IEEE LTSC, IMS, ADL, AICC and some other European groups are developed these standards. In particularly, AICC is working on an independent and unique industry, ADL works on generalizing and IEEE LTSC are looking to create a formal standard.

6.3 T-SCORM standard

For increasing SCORM standard benefits, T-SCORM is extend of SCORM to improve it in searching and navigation making LOs available via iDTV (Digital Television) platform. One of the benefit extension is to enable a system which made for T-Learning to search information based on LOs. **T-SCORM** extension is an advanced of metadata information from SCORM standard based on LOM standard. So, it should be proposed the addition of new elements giving more emphasis on the metadata information regarding to iDTV. These new elements in the LOM structure, identify specific information for iDTV (interactivity level, copyright, description on content in digital format, etc.).

6.4 IEEE (Learning Technology Standard Committee) LTSC standard

This standard is one of the most IEEE publications. It is characterized by developing LOM (Learning Object Metadata). It is recommended guidelines for educational and training systems, especially software components, tools, technology solutions that enable development and maintenance. However, it does not support details about implementation of specific technologies, the standard represents a high-level model of e-learning system

architecture throughout standardize five categories: General items; Learn more about; Content; Data and metadata, and Management systems and applications.

6.5 IMS Standard

IMS project is national infrastructure of higher education in the United States. This project is managed by the union called EduCAUSE or Educom in forms of hundreds of universities and educational institutions. The project aims to establish standards for dealing with problems associated with the increasing use of new technologies in teaching and learning.

IMS standard provides the following sections in this regard [12]:

- **IMS Learning Resource Meta-Data Specification:** Describes learning resources in order to search
- **IMS Enterprise Specification:** It is used for sharing data about learners, courses etc.
- **IMS Content Packaging Specification:** It creates and shares content objects, reusable learning content
- **IMS Question & Test Specification:** It shares test items and other assessment tools
- **IMS Learner Information Package Specification:** It organizes information so that the learner can learn the system for a specific user and can provide appropriate needed responses.
- **IMS Reusable Competency Definition Specification:** It is used for the descriptions, references and related transactions with key characteristics of a potential
- **IMS Simple Sequencing Specification:** It illustrates how learning is arranged and provided in a specific sequence for a learner.
- **IMS Accessibility Specification:** This section provides guidance for other sectors and it aims to ensure the availability and ease of use of the standard specification.
- **IMS Learning Design Specification:** This section is used to introduce the interaction scenarios for creator of subject and educational courses.
- **IMS Digital Repositories Specification:** It integrates online learning system with other data sources.

6.6 The Learning Object Management (LOM)

This standard is formal standardization licensed by the ISO/IEC SC36 with one option being to “fast-track” the standard through a high-level JTC1 committee. As a further result, LOM is subjected as standard issued by the IEEE LTSC and the SC36 subcommittee to work together in the

future to develop a “next generation” of this metadata standard. This version proves that is much closer in orientation to the “minimalist” Dublin Core approach to metadata than to the technically demanding, “structuralist” approach represented by the LOM.

7 E-LEARNING QUALITY STANDARDS

Quality is a key of learning success in general. The following observations are considered among the most important E-learning quality issues [13,14]:

- Learner orientation;
- Developing quality in learning process;
- Quality must be a key role in education policy;
- Quality services should be considered;
- Quality standards should be implemented.

Following part browses the most common E-learning quality standards:

- **ISO/IEC 19796-1**

It is published in 2005, and aimed to develop and improve quality systems in the educational processes, activities and services. The standard is used as a reference to support adaptation specific requirements of the organization.

Since 2007, this standard became a reference model and adapted the needs of organizations. In 2012, the official released of an international quality standard for e-Learning programs Open ECB-Check (e-Learning in Capacity Building) supports in measuring the success of E-learning programs and allows development. ECB Check supports a set of an E-learning quality criteria that helping in design, development, management, delivery and evaluation program, as well as the quality of learning materials, methodology, media, technology and e-tutoring.

- **ISO 9126**

This quality standard proposed a guideline to evaluate the e-Learning systems for teachers and educational organizations. The aim is to support decision making regarding evaluating the quality of existing systems and also to develop educational systems by increasing the usability by adding quality attributes such as: consistency, simplicity, legibility, and user satisfaction as a global characteristic of the model. In 2010 the ISO 9126 model is used for selecting as standard quality for evaluating course management system in fields of design, develop and deliver e-Learning content and measure the E-learning outcomes. Furthermore, ISO 9126 model is customized to identify acceptance criteria and evaluate a B2B (Business to Business) applications by adding additional characteristics to existing quality models. Besides that, the ISO 9126 is used for evaluating the mobile learning by

adding the following characteristics: Metaphor, interactivity, learning content.

8 E-LEARNING COMMUNITIES STANDARDS

This section presents an overview of the major organizations that contribute to the development of e-learning standards (e.g. IMS, IEEE LTSC, and the ISO/IEC). [15,16]

- **IMS Global Learning Consortium, Inc. (IMS)**

IMS is a consortium that develops promotes open specifications for facilitating online distributed learning activities that called E-learning. IMS is the only organization developed standards for school representation in K-12 sector. This representation includes governmental representation from education ministries. Contributing members are able to vote on the IMS technical board for the acceptance, rejection or revision of specification drafts, then envision for standards.

- **Aviation Industry CBT Committee—AICC Standard**

AICC standard was primarily formed on the need of standardizing of computer training for using in airline industry, but now, it is used for reusability, interoperability in online learning and applications such as health care, financial services, higher education and telecommunication. AICC adopted all computer-based training and include supplying, controlling, delivering and monitoring the results of management systems training and internet courses [16]. AICC issued three types of documents, they are:

- AICC guidelines and recommendations.
- AICC reports and technical articles
- AICC working documents.

- **Dublin Core Metadata Initiative DCMI**

This organization is adopting interoperability metadata standard, especially metadata vocabulary.

Dublin Core defines metadata that developed i.e. title, creator, subject, description, publisher... etc. This standard represents XML and RDF languages. There are some available documents issued by Dublin Core: Dublin Core Template, MyMeta-Maker, Reggie-The Metadata Editor, DC-dot [17].

- **Ariadne Foundation**

This is a non-profit association and concerned with area of metadata. The foundation grouped metadata into six categories: General information, Semantics, Pedagogical, technical, indexation, annotations.

- **Advanced Distributed Learning Initiative ADL**
ADL Initiative is established to provide U.S Department of Defense and white House Office

Table 2. Examples of E-learning trends standards.

Trend	Standard name
Architectures	IMS Guidelines for Developing Accessible Learning Applications, IEEE Learning Technology Systems Architecture, Open Service Interface Definitions
Digital Repositories	CWA 15454 Simple Query Interface, IMS Digital Repositories Interoperability
Content Aggregation	ADL Content Aggregation Model (CAM), IMS Content Packaging (CP), IMS Simple Sequencing (SS), IMS Common Cartridge
Metadata	IMS Learning Resource Metadata Information Model, IEEE LOM, Dublin Core Metadata Element Set (ISO 15836), Metadata for Learning Resources (ISO 19788), Dublin Core interoperability
Accessibility	IMS Learner Information Package Accessibility, ISO/IEC 24751
Competency Definitions	IMS Reusable Definition of Competency, IEEE Data Model for Reusable Competency Definitions
Quality Assessment	ISO/IEC 19796 IMS Question and Test Interoperability (QTI)
Vocabularies	ISO/IEC 2382, AICC glossaries, IMS Vocabulary Definition Exchange
Runtime	ADL SCORM Run-Time Environment, AICC/CMI Guidelines for Interoperability, IMS Shareable State Persistence

for Science and Technology with development plans for standardization for learning researches. The main objective of ADL is to support a high quality education and training tailored to learner's preferences, highly cost-effectively and expand accessibility. Another main objective is to enrich SCORM standard in order to be compatible with other system in interoperability functionality and encapsulate to its content. There are several international organizations working on standardizing E-learning technologies. Each develops different learning concepts technology standards which consist of a set of definitions, specifications, guidelines and recommendations. Table (2) browses example of E-learning standards trends based on citation and compatibility with LMSs [18].

9 CONCLUSION

This research gives a summary of using standards in e-learning industry, browses the most cited organizations that authorized to approve

specifications as standards such as SCORM, Metadata, Interoperability, AICC, IEEE LTSC and IMS. Also it presents the quality standards and its frameworks, gives an overview on personalization and adaptation learning. On the other hand, points on the trends of e-learning standards.

REFERENCES

- [1] Norm Friesen, CanCore Initiative, "Interoperability and Learning Objects: An Overview of E-Learning Standardization", *Interdisciplinary Journal of Knowledge and Learning Objects*, Vol1, 2005.
- [2] Ileana Adina UȚĂ, "E-learning Standards", *Informatica Economică*, nr. 1 (41)/2007.
- [3] Mohamed A. Khamis "Adaptive e-Learning Environment Systems and Technologies", The First International Conference of the Faculty of Education, "Education Future Prospectives", Albaha University, during the period 13-15 / 4/2015.
- [4] Aimad Qazdar1, Chihab Cherkaoui2, Brahim Er-Raha3, Driss Mammass4, "AeLF: Mixing Adaptive Learning System with Learning Management System", *International Journal of Computer Applications* (0975 8887), Volume 119 - No. 15, June 2015.
- [5] Hazem M. El-Bakry*, Ahmed A. Saleh, "Adaptive E-Learning Based on Learner's Styles", *Buletin Teknik Elektro dan Informatika (Bulletin of Electrical Engineering and Informatics)*, Vol. 2, No. 4, December 2013, pp. 240-251, ISSN: 2089-319.
- [6] Advanced Distributed Learning (ADL) Collaboratories, "Choosing a Learning Management System Advanced", Version 4.10, 25 Sep, 2015.
- [7] Francisco M. Silva, Aquiles M. Filgueira Burlamaqui, Karla R. do Amaral Demoly, João Ph. de Freitas Pinto, "Providing an Extension of the SCORM Standard to Support the Educational Contents Project for t-Learning", *Creative Education*, 2015, 6, 1201-1223, Published Online June 2015 in *SciRes*.<http://www.scirp.org/journal/ce> <http://dx.doi.org/10.4236/ce.2015.611118>.
- [8] J.L. Fernández, J.M. Carrillo, J. Nicolás, M.I. Carrión, "Trends in e-Learning Standards", *Proceedings published by International Journal of Computer Applications® (IJCA)*, 2011.
- [9] A.M. Bianco, M. De Marsico, M. Temprini, "Standards for e-learning", *QUIS- Quality, Interoprability and standards for e-learning*, 2004-3538/001-001 ELE- ELEB14.
- [10] Essaid El Bachari, El Hassan Abdelwahed and Mohammed El Adnani, "E-Learning Personalization Based On Dynamic Learners' Preference", *International Journal of Computer Science & Information Tech (IJCSIT)*, Vol 3, No 3, June 2011.
- [11] Essaid El Bachari, El Hassan Abdelwahed, Mohamed El Adnani, "Design of An Adaptive E-Learning Model Based On Learner's Personality", *Ubiquitous Computing and Communication Journal* Volume 5, Number 3.
- [12] R. Sivakami, G. Anna Poorani, "SCORM/AICC Compliance in Learning Management System and e-Learning: A Survey", *International Journal of*

- Engineering And Computer Science ISSN:2319-7242 Volume 4 Issue 6 June 2015, Page No. 12894-12897.
- [13] Ana M. Mogoš, "The Use Of Quality Management Systems For E-Learning", *The Sixth International Conference on e-Learning (eLearning-2015)*, 24- 25 September 2015.
- [14] Chirag Indravadanbhai Patel, "A survey paper on e-learning based learning management Systems (LMS)", *International Journal of Scientific & Engineering Research*, Volume 4, Issue 6, June-2013, ISSN 2229-5518.
- [15] Ayan Roy*, Kaustuvi Basu, "A Comparative Study of Statistical Learning and Adaptive Learning", *International Journal of Advanced Computer Research* ISSN (Print): 2249-7277 ISSN (Online): 2277-7970 Volume-5 Issue-21, December-2015.
- [16] Karan Venupure, Sonal Chitare, Chandra Rangan, Meghana Lokhande, "Survey on Personalized E-learning System", *International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization)*, Vol. 3, Issue 11, Nov, 2015.
- [17] Marija Blagojević, Živadin Micić, Danijela Milošević, "Development of Standards In E-Learning", *The Sixth International Conference on e-Learning (eLearning-2015)*, 24-25, September 2015.
- [18] Lamia Mahnane, Mohamed Tayeb Laskri and Philippe Trigano, "A Model of Adaptive e-learning Hypermedia System based on Thinking and Learning Styles", *International Journal of Multimedia and Ubiquitous Engineering* Vol. 8, No. 3, May, 2013.

Cross-layer quality of service protocols for wireless multimedia sensor networks

Ablah AlAmri

Jeddah Community College JCC, King Abdulaziz University KAU, Jeddah, Saudi Arabia

Manal Abdullah

Faculty of Computing and Information Technology FCIT, King Abdulaziz University KAU, Jeddah, Saudi Arabia

ABSTRACT: Wireless Multimedia Sensor Networks (WMSNs) recently gained attention of many researchers, WMSNs have sensor nodes that are deployed to extract information from surrounding environment, processing information locally then transmit it to sink/base station wirelessly. Multimedia data include audio, image, and video. Sensor nodes are limited in battery, memory and CPU capability. Multimedia data is larger in volume than scalar data, thus transmitting multimedia data require stick constraint on quality of services in term of energy, throughput and end to end delay. Cross layer architecture is a new concept which combine several layers to allow integration and exchange information among them more efficiently than classical layered approach. This paper discusses and compares between existing cross layer protocols for WMSNs which cross functionalities of adjacent or non-adjacent layers.

1 INTRODUCTION

Recently with advanced technology in Micro Electro-Mechanical Systems (MEMS), communication through wireless and electronic digital, developing a sensor nodes with low power, cost and provide different functions are become feasible [1]. Wireless Sensor Networks (WSNs) is a network that is created by deploying a large number of these sensor nodes. Sensor nodes have ability to sense the environment and send scalar data to the sink/Base Station (BS) [1–3].

Many applications such as monitoring and surveillance required sensing the environment and send multimedia data not only scalar data. Multimedia data may include video, sound or even image [3]. More sophisticated sensor nodes can be made by integrates a cheap component such as CMOS cameras and microphones to the sensor nodes to create a Wireless Multimedia Sensor Networks (WMSNs). WMSNs have sensor nodes that are capable of capture and communicate streams of multimedia data over a wireless channel to the base station [4, 5]. Sensor nodes have limitations in terms of power provider, processing capability, and storage memory [1–3, 5].

WMSNs have many challenges due to the type of transmission data, transition medium and limitation of sensor nodes. WMSNs required

Quality of Service (QoS) in terms of increasing throughput and reduce end to end delay. The packets of multimedia data is very sensitive to delay and losses, because losing packets or arriving after deadline leads to distortion in received multimedia data [6]. Also reduce energy consumptions are important in order to increase the network lifetime [7].

Layered architecture such as TCP/IP consists of five layers (application layer, transport layer, network layer, data Link layer and physical layer) [1]. Layered architecture is independent, every layer is separated and encapsulate from each other, and only adjacent layers can communicate directly [7]. A new architecture that combines several layers to allow integration and exchange information among them more efficiently than layered approach, name it cross layer architecture [7, 8]. Layers in this architecture are dependent and can share variables between non adjacent layers [7]. More efficient cross layer protocols will improved transmission performance and satisfy the stringent quality of service required for multimedia transmission in WMSNs [8]. Cross layer architecture is gained the attention of many researchers recently to produced different cross layer protocols for WMSNs. The paper focused mainly on a cross layer protocols. It discusses, categorizes and compares between existing cross layer protocols for WMSNs that join adjacent or non-adjacent layers without including physical layer.

The rest of the paper is organized as follows: in section 2, a brief background about WMSs and WMSNs, section 3 discuss existing cross layer WMSNs protocols, and conclude the paper at section 4.

2 BACKGROUND

The advanced technology in low power circuits, cheap sensor nodes with different functions open the door to the sensor networks, thousands number of sensor nodes are deployed to cover specific area, these deployed sensor nodes cooperate together to create sensor networks [1].

2.1 *Wireless Sensor Networks (WSNs)*

The sensor nodes which have ability to sensing the environment, processing the data locally, and communicating giving a birth of Wireless Sensor Networks (WSNs) by corporate the effort of a large number of deployed sensor nodes [1].

Wireless Sensor Networks (WSNs) have sensor nodes that deployed in a physical environment, these nodes are capable of capture the events and communicate streams of scalar data over wireless channel to the base station (sink node), sensor nodes responsible for the fusion and should process the data locally and only transmit the required data. Sensor nodes send the data via multi-hop through the sink, these nodes are usually scattered in a sensor field. Sensor node may transmit their packets directly to the sink through single hope path, or send it to another node in order to forward it to the sink through multi hop path. The sink may use Internet or wireless network such as Wi-Fi to communicate with the end-user. WSNs nodes are limited in resources such as battery, memory and CPU capability [1].

2.2 *Wireless Multimedia Sensor Networks (WMSNs)*

WMSNs are a new branch of wireless sensor networks. The integration of inexpensive components such as CMOS cameras and microphones to sensor nodes giving a birth of a new branch of wireless sensor networks namely Wireless Multimedia Sensor Networks (WMSNs). These new nodes are smart device which have the ability to capture and transmit multimedia data such as video, audio and even image to the base station [4].

The difference between these two networks (WSNs and WMSNs) is due to the type of data that transmit through wireless channel and design constrains in WMSNs. The design of WMSNs is focused on reducing the end to end latency and

speed the delivery of multimedia data packets to the destination, because multimedia data packet is very sensitive to the delay and losses, losing these packets or arriving after deadline leads to distortion in received multimedia data [6].

WMSNs begin used in many applications such as: multimedia surveillance sensor networks to track the object and take appropriate actions [9]. Track the missing persons and locate their places, also used to identify the criminals, thieves or potential terrorists, control systems to monitor the traffic to avoid congestion, also used in many activities such as thefts, road accidents, traffic violations [4]. In smart homes for energy efficient; control heating, cooling, and light system based on human activities. Advanced health care delivery; remote medical center monitor patients parameters to infer any emergency situations, patients carry medical sensors to detect their body parameters such as temperature, breathing, and pressure. Use WMSN sensors to monitor the environment and civilian structure such as bridges. Also WMSN is used sensors to control industrial process [4, 10]. All these applications require a Quality of Service (QoS) for multimedia transmission [9].

2.3 *Sensor networks layered and cross layer architecture*

Layered architecture is hierarchy, layers are independent, separated, and encapsulate from each other, direct communicate allowed between only adjacent layers [7]. TCP/IP model is example of this architecture, it is consists of five layers: application, transport, network, data link and physical layers

These five layers are separated and only adjacent layers can communicate [1]. In addition, the sensor nodes have different planes that help the sensor nodes to coordinate the task and save energy. The power planes to monitor the sensor node power, mobility planes to monitor the movement, and task management plane distribute the task among the sensor nodes. Figure 1 shows layers protocols stack.

The communication protocol plays a major role for correct functionality for these networks, limited resources and wireless communication medium prevent from using the traditional layered architecture in WSNs and WMSNs. For that cross layer architecture is produced. Cross layer architecture is a new design which combines several layers—adjacent or none adjacent—to allow integration and exchange information among them more efficiently than classical layer, more efficient cross layer protocols will satisfy the stringent quality of service required for multimedia transmission in WMSNs [8].

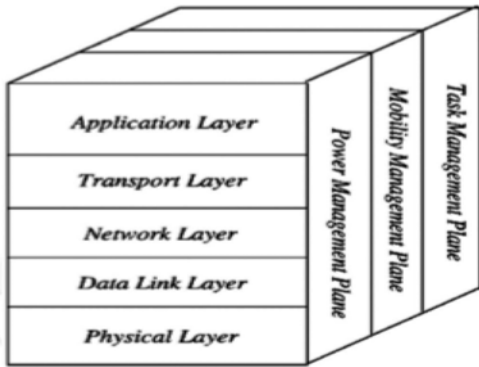


Figure 1. Layers protocols stack [1].

Many researchers provide different types of cross layer architecture, some researchers combine adjacent layers such as: network and MAC layers protocols proposed in [6, 11], application and transport layers [12], transport and network layers [13]. Other researchers combine none adjacent layers such as application and network [14]. Other researchers combine more than two layers [15] where transport, network, and MAC layer are joint. Researchers also defined performance metrics and used network simulation to evaluate their protocols as describe in next sections.

3 CROSS LAYERS FOR WMSNs PROTOCOLS

There are many researchers produced different designs and protocols for cross layer architecture to increase the data gathering from WMSNs nodes to the base station, reduce the latency, increase the bandwidth and reduce the energy consumptions. This is to show how WMSNs can be more efficient depending on the constraints and the requirements of QoS on specific application. This section summarizes what has been done on cross layer protocols for Wireless Multimedia Sensor Networks (WMSNs). We can classify WMSNs protocols based on the routing techniques into three categories as shown in Figure 2, these categories are describing below:

3.1 Multichannel routing

Multichannel routing divided the bandwidth into different separate channels and dedicate each channel for specific packets types [15].

A protocol proposed by Fard et al. in [16, 17] is a cross layer multichannel QoS-MAC protocol supposed a clustered network using any existing

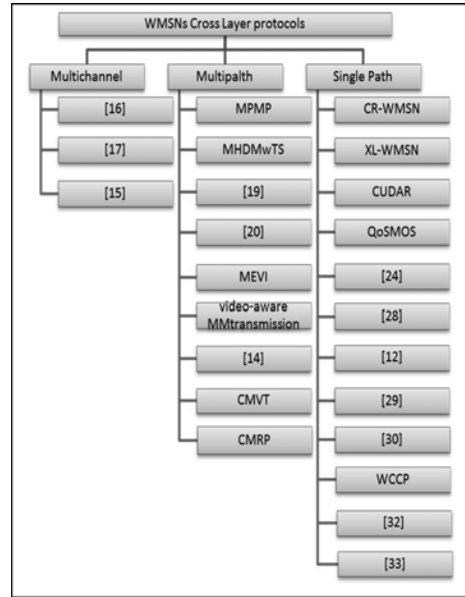


Figure 2. WMSNs cross layer protocols.

clustering techniques, and supposed in each cluster three types of node: active nodes defined as nodes need to send data, passive node defined as nodes did not have any data to send, and cluster head.

There are three channels where all sensor nodes can transmit or receive through these channels, in initially stage of network deployment, channel 1 is assigned to the cluster head, the cluster head dynamically assigned channel 2 to sensor nodes, channel R is shared between all sensor nodes.

The MAC protocol proposed is lie in three phases: A. request phase where the active nodes send to the head a request message (REQ), this message contains QoS requirements, B. Scheduling phase where the cluster heads broadcast the scheduling messages to the active nodes, these messages contain the channel and time slot for each node, the scheduling is based on the QoS requirements and the priority, multimedia assigned to highest priority, C. Transmission phase where an active node start to send its data on assigned channel to the cluster head and receive acknowledgement messages (ACK), then the cluster head classifies traffics based on its priority where the video traffic assigned to highest priority, finally cluster head schedules for sending the data to the sink.

A multichannel cross-layer architecture produced in [15] by Çevik, T. and A.H. Zaim, that combined three layers; transport, network and MAC layers. The architecture maximize the network life time by fairly distributed the load among all nodes. Classify the incoming packets to three types: routing packet

request that reserve the sources, this packet has highest priority. Real-time packet for unusual event, this packet has less priority than previous one. Non real time packet for delay-tolerant data, this packet has the least priority. The architecture contains three schedulers. Scheduler 1 for classifying the incoming packets to appropriate queues. Scheduler 3 for the queues that contain real-time packet, it used round robin to pull the packets. Scheduler 2 for Request queue first then for non-real time. In WMSN application such as surveillance when unusual event detected then a stream of multimedia data should be transmitted to the sink, if this stream transmitted via single path then the energy of the nodes that rely on this path is depleted. To balance the load, segment the original stream into flows. The number of the flows is defined depending on the number of the paths available that satisfy the QoS required. The multichannel structure is used by dividing the bandwidth into separate channels, dedicate one channel for non-real time data and control messages, the rest channel for real time data. Create a rout request message for each flow, send this request via control channel, used routing algorithm to select the next hop. Reserve the resource during construct the path. The path is discovered using a modified AODV, the next hop is defined based on load balanced routing algorithm, where consider the hop count as QoS parameter.

3.2 Multipath routing

Multipath routing, this type of routing during route discover stage discovered different separate path from the source node to the sink [13].

Protocols in [13, 18] combine transport and network layers, [19–22] combine application, network, and MAC layers, [14, 23] combine application and network layers.

Earlier protocols that combine a cross layer concept with context awareness is proposed by [18]. The aim of this protocol is to maximum the gathering important information instead of maximum throughput. Important information or information value based on the application, for example in ocean monitor, the sound information is more important than video. But in fire monitor the video is more important. The protocol split video streams from audio streams. The protocol is Multi-Path Multi-Priority transmission (MPMP) protocol, At network, used Two-Phase geographic Greedy Forwarding (TPGF) to discover the large number of paths from source node to base node and delay for each path, then at transport layer a Context Aware Multi-path Selection algorithm (CAMS) is used to select the large number of disjoint paths, There is no limited in number of selected paths as in [13]. The most important stream assigned to higher

priority and to the best routing path that guarantee minimum end to end delay.

A Minimum Hop Disjoint Multipath routing algorithm with Time Slice load balancing congestion control scheme (MHDmWTS) is produced by Sun, G., et al. [13]. They used a minimum hop to reduce the delay and increase the reliability in WMSN. Minimum Hop Disjoint Multipath routing algorithm (MHDm) is built up three disjoint paths: primary path, alternate path and backup path. This algorithm has two phases: path build up and path acknowledgment. In path build up, when source sensor node activated, then it sends request package to build up a path to the neighbors that has hop count smaller than the source. The primary path is the first package reach the sink, the primary path is the least delay. The packages continue come to the sink from different routes. Compare every new package with the primary path and discard it if there is a joint node else the alternate path is build up. At path acknowledgment phase, sink send to the source acknowledgment message (ACK). ACK contains path and time information. To balance the multiple path load use time slice load balancing congestion control scheme. When sink node allocate time called time slice for the path, primary path should take more time than others. After the source node receive the ACK from the sink, it starts to transmit the data, each path take a time slice if it is up then the sink switching the transmit data to another one. MHDmWTS protocol reduces the end to end delay, controls and prevents the congestion.

Produced protocols in [19][20] consider constraints such as bandwidth, end to end delay and reliability. This architecture consist of multiple components:

Traffic classifier module to classify the types of frames, application layer encapsulates frame type, frame priority, and group of pictures size (GOP size) to the header and send the frame to the route classifier module. Then route classifier module find three disjoint paths reach the QoS requirements used multipath routing algorithm. The source increases the GOP size when there is none of available bandwidth can provide bandwidth required. MAC layer used prioritized scheduling to access the medium.

Two protocols for heterogeneous networks are produced in [21][22] names as MEVI and video-aware MMtransmission where Camera Nodes (CN) are the cluster heads and Sensor Nodes (SN) are members. At network a multi-hop hierarchical routing is used: the intra-cluster communication between the members and its Cluster Head (CH) follow TDMA schedule. And the inter-cluster communications between CHs and base stations, create disjoint path routes to the base station, then classify these route based on residual energy, hop count and link quality. At application layer classify the frame into I-frame, P-frames and B frames.

Where I high priority frame and other two type less priority. High priority frames assigned to best priority paths. The aims of video-aware MMtransmission protocol are to balance the load and enhance the video quality. Bae et al. [14] cross layer QoS architecture. Used at application layer packet marking algorithm to mark each traffic based on its priority. Used at network layer, multipath algorithm to classify the packets into different color such as green, red and yellow in order to distribute the packets into different path. The shortest path is assigned to the green packets that required high quality transmission, assigned the alternating path to red packets that consider the level of energy, assigned yellow packets to the path consider the level of quality and distance to the sink. Moreover the authors suggest routing table that keep only information about the nodes that in the path to the sink. In order to avoid communication overhead, the architecture stored the important information from each layer at shared database.

Cross-layer and Multipath based Video Transmission scheme (CMVT) is protocol produced by Guo et al. in [23], at application layer used MPEG-4 as in [24], where encode the video and marked the frame with video type. At network: two main components route discovery where discover all possible rout from source to the sink, and data transmission where evaluate all the paths and categorize them into three categories based on evaluate value, assigned the frame to the path based on frame type and priority, where higher priority frame is assigned to best path.

A Cross-Layer-Based Clustered Multipath Routing CMRP is protocol proposed in [25] by Almal-kawi et al. for heterogeneous networks. CMRP combine network and MAC layers, CMRP defined two thresholds: upper and lower thresholds, used the upper select the cluster heads of 1st level and node members, then establish a link between all cluster heads used the lower thresholds, the 2nd level cluster head will selected toward the base station, and create a multipath routing, sort these paths based on different criteria such as delay, hop counts, bandwidth, and link quality. Reserved the best paths for multimedia data, and the other paths for other data type. At MAC layer used TDMA with time slot, and give a high priority for multimedia.

3.3 Single path routing

Single path routing discovers a single path from the source node to the sink at route discover stage a single path [11].

Different proposed protocols with different combine layers, [6, 11, 26, 27] provide interaction between network and Mac layer, [24, 28] provide interaction between application and MAC layers.

Two protocols proposed by [11] and [6] namely XL-WMSN and CR-WMSN to decreases the end-to-end packet latency and increases the throughput of multimedia traffic, but both of them not enhance the network lifetime. Both protocols consisting of multiple components working together to meet the QoS that is required by multimedia applications. They use admission control scheme that eliminate nodes with less remaining energy during rout discovery, and participate to the routing path only nodes with enough energy.

To select the appropriate path from source to destination with least delay, authors in [11] [6] used average packet service time PST_{avg} which is a sum of all possible delay such as queuing delay, network and MAC layer delay, and transmission delay to provide information about node load. Both protocols used channel utilization "Utili" which is a proposed mechanism to check the channel regularly to give indicator of local contention. Also both used reactive approach to establish the path and used hop count to avoid long path. Moreover in [6], it has restricted on the path length to be close to shortest path. Moreover, authors in [6] classified incoming traffic into three classes based on its priority, where gives video traffic highest priority. Static duty cycle where nodes have a fixed interval of sleep and awake to listen to the medium is energy consuming and not suitable for WMSN, authors in [6] produced dynamic Duty Cycle Assignment (DCA) based on traffic type, where nodes have mainly video traffic expected to have longer duty cycle and consume more energy than nodes have other type of classes. An extended of CR-WMSN protocol proposed by [26] name as CUDAR. The CUDAR aims are to guarantee a high throughput, low delay, jitter and set up time. At application layer used delay and channel aware routing, at MAC layer used a modified CSMA/CA.

A cross-layer QoS architecture (QoS MOS) that developed a Cross-Layer Communication Protocol (XLCP) proposed by Demir et al. [27], this protocol reduce delay and enhance throughput. XLCP contains different elements, classifier that classifies the packets and mark packets that require QoS, then use scheduler to broadcast a Request to Send Investigation (RTS-I), and the sensor nodes that have enough battery and close to the sink node will reply and investigate to the route.

A cross layer optimization method is proposed by Zhang, J. and J. Ding. in [24], it proposed mapping algorithm that use MPEG-4 standard to define the video frames, it classifies the video frame to three types; I-frame is the most important type and has highest priority, P-frame is less important and has less priority than previous one, B-frame is the least important and has the least priority among them. When needed, the medium IEEE 802.11s is used, IEEE 802.11s use Enhanced Distributed Channel Access

(EDCA) to support QoS, EDCA categories accessing medium to four Access Categories (AC) based on the traffic type: background (ACO), best effort traffic (AC1), video (AC2) and audio (AC3). The mapping algorithm is mapping the packets to Access Categories (AC) based on traffic load using a threshold.

A cross layer QoS architecture is produced by [28] to provide QoS for urgent real time traffic for emergency situation. Proposed two tiered service differentiation mechanism TTSDM at MAC layer that classify the traffic to: urgent and non-urgent traffic then divided each one of these traffic to real time—for multimedia traffic-, non-real-time traffic and best effort traffic. There is a predefined threshold, if urgent real time queue exceed the threshold then MAC layer interact with application layer to lower the data rate, otherwise MAC layer interact with application layer to increase the data rate. Authors used Data Rate Adjustment Scheme (DRAS) at application layer to control the data rate.

Paniga, S., et al. presented a cross layer architecture [12] combines application and transport layer. At application layer, they used a hybrid DPCM/DCT coding algorithm. It is a predictive compression scheme that provides acceptable compression and low the complexity. Using I frame with high priority and P frames in predictive coding, in case of loss frame: only frames with high priority are retransmitted. At transport layer congestion control mechanism is used with two thresholds, stop threshold and restart threshold at the buffer. This protocol used a static routing at network layer, IEEE 802.15.4 CSMA protocol with Clear Channel Assessment (CCA) at MAC layer.

Produced a cross layer solution to maximize the network lifetime in [29], You et al. divide network lifetime problem into sub problems in different layers, where each layer solved part of this problem. At application layer use Pairwise Distributed Source Coding (Pairwise DSC) to collect information about neighbor's nodes, to avoid the redundancy use Slepian-Wolf Distributed Source Coding (DSC). Transport layer used source rate adaptation and at network solved routing problem.

An integration between transport and MAC layer produced in [30] by Tiglaio et al, where two mechanisms are provided. Using Negative Acknowledgment (NACK) based Repair Mechanism, the receivers and intermediate node have ability to detect packet losing and send back Repair Negative Acknowledgment (RNACK), if the intermediate nodes have the losing packets in their cache then retransmit it to the destination, else propagated RNACK to the source.

A WMSN Congestion Control Protocol (WCCP) proposed by in [31] by Aghdam et al. for control the congestion and received high quality video. WCCP provide interaction between Application, transport, and MAC layers. Application

layer defined frame type and packet number. Transport layer: WMSN Congestion Control Protocol (WCCP) is two parts protocol where a Source Congestion Avoidance Protocol (SCAP) at the source nodes, and a Receiver Congestion Control Protocol (RCCP) in the intermediate nodes. SCAP used a GOP size prediction method to prediction the Congestion, and used in case of congestion drop the less important frame and keep the I-Frame to improve the video quality. RCCP proposed a congestion control to detect or predict the congestion and send notification if congestion is detected. WCCP is control the congestion and received high quality video. At MAC layer used IEEE 802.15.4 protocols to transmit the data. S. Sridevi and M. Usha proposed a cross layer framework for heterogeneous WMSNs in [32]. This framework provides interaction between MAC, network and transport layers, Classify the traffic into different classes with different priority. Cluster the sensor nodes to different clusters, each cluster has cluster head. When source node need to send a data, it send the data toward the cluster head using TDMA schedule, authors used a dynamic priority for sensitive data and allocate more slots for sensitive data. Used congestion detection mechanism based on the total number of packets in the queues (Qcw). If the (Qcw) is less than the threshold then the network loaded is normally. The author did not provide a simulation for this framework.

Unlike traditional geographic routing algorithms that selected the next node based on short distance between node and destination, and ignore the link quality. A cross layer QoS protocols is produced by Duan et al. in [33] that joint transport and network layers. The authors produced geographic routing metrics based on link quality and shortest distance to the destination. Link quality evaluated by used number of Hello packets that received to the node from its neighbors. Bit Error Rate (BER), payload length, and wireless environment are parameters affect the packet loss rate. Unlike the traditional where the payload length is fixed, this lead to poor video quality because of the dependability between the BER and the packet loss rate. The author's scheme change the payload length depends on transmission quality feedback and hop count, and this broke the dependability between the BER and the packet loss rate, and maintain low level of the packet loss rate when the BER is increase. Authors also encode the video packet using short-length LT code technique before transmit the video packets. In case of lost packets use these encoded packets to recover the data. This protocol reduces the decoding overhead and the packet loss rate, enhance the video quality and the network efficiency.

A summary of all discussed cross layer protocols for WMSNs shown in Table 1.

Table 1. A summary of all discuss WMSNs cross layer protocols.

Protocols	Aims	Cluster	Classify traffic	Energy-aware	Route discovery	Joint layers	Application	Transport	Network	MAC
Multichannel routing										
[16] [17]	Energy-efficiency, throughput, and data reliability	Yes	Yes	–	–	Network + MAC layers	–	–	Multichannel	Modified 802.11
[15]	Improved energy efficiency and delay	–	Yes	Yes	–	Transport + network + MAC layers	–	Segmentation of the stream into flows	AODV based routing algorithm	Three Schedulers
Multipath routing										
MPMP [18]	Maximum valuable information	No	Split the video and audio stream	No	–	Transport+ network layers	–	CAMS	TPGF	–
MHDMwTS [13]	Reduce delay Prevent congestion	No	No	Yes	–	–	–	Congestion control scheme	MHDM	–
[19] [20]	Maximize number of video sources	No No	Yes Yes	– –	– –	Application + network + MAC layers	Distributed source coding	– –	SDMR	802.11e multi-rate transmission mode
MEVI [21]	Enhance, network lifetime, scalability and reliability	Yes	Yes	Yes	–	–	–	–	Multi-hop hierarchical routing protocol	TDMA
Video-aware MMtransmission [22]	Balance the load and enhance the video quality	Yes	Yes	Yes	–	–	–	–	–	–
[14]	Improved the transition rate, packets loss and end to end delay.	No	Yes	–	Modified routing table	Application + network layers	Packet marking algorithm	–	Multipath algorithm	–
CMVT [23]	Increase network lifetime, enhance the video quality and reduced channel conflict	No	Yes	Yes	–	Application + network layers	MPEG-4	–	Route discovery + data transmission function.	–

(Continued)

Table 1. (Continued)

Protocols	Aims	Cluster	Classify traffic	Energy-aware	Route discovery	Joint layers	Application	Transport	Network	MAC
CMRP [25]	Enhance the reliable, throughput and energy efficient	Yes	Yes	–	–	Network + MAC layers	–	–	Cluster-based multipath routing	TDMA with time slot
Single Path Routing										
CR-WMSN [1]	Minimize end to end delay	No	No	Yes	Reactive approach	Network + MAC layers	–	–	Delay and channel aware routing	Modified 802.11
XL-WMSN [2]			Yes							DCA
CUDAR [3]	High throughput, low delay, jitter and set up time.		–							Modified CSMA/CA
QoS MOS [4]	Reduce delay, enhance throughput and reliability		Yes	Yes	–				Geographic routing mechanism based on location awareness	CSMA/CA-like
[24]	Forward video sequences	No	Yes	–	–	Application + MAC layers	MPEG-4 standard	–	–	IEEE 802.11s
[28]	Lower delay for emergency situation	Yes	Yes	–	–		DRAS	–	–	TTSDM
[12]	Effective multi-hop streaming video	No	Yes	–		Application + transport layers	DPCM/DCT coding scheme	Congestion control mechanism	Static routing	IEEE 802.15.4 CSMA with (CCA)]
[29]	Maximize the network lifetime	No	No	–	–	Application + Transport + Network Layers	Pairwise DSC	Source rate adaptation	Overall link rate control	–

[30]	Improved the performance and efficient the energy.	No	No	-	-	Transport + MAC layers	NACK-based repair mechanism	Adaptive retransmission mechanism
WCCP [31]	Control the congestion and received high quality video. Reduce delay	No	Yes	-	-	Application + transport + MAC layers	SCAP + RCAP	IEEE 802.15.4
[32]	Reduce delay	Yes	Yes	-	-	Transport + network + MAC layers	Congestion detection scheme	TDMA slot assignment
[33]	Reduce the decoding overhead and the packet loss rate, enhance the video quality and the network efficiency	No	No	The geographic routing	-	Transport + Network Layers	The short-length Luby transform (LT)	The hierarchical routing protocols The geographic routing

4 CONCLUSION

The advanced technology in low power circuits, inexpensive CMOS cameras and microphones gave a birth to WMSNs. WMSNs are useful in many applications especially in surveillance to track the object and take appropriate actions, traffic avoidance and control systems to monitor the traffic to avoid congestion, advanced health care delivery and smart homes. All these application required to transmit video, audio or even image, transmit multimedia data required constraint on QoS that define by the application. Architecture such as cross layer where boundaries between layers are eliminating, this is allowing exchange parameters between layers more efficiently to increase the performance. Across layer protocol will satisfy the QoS required for multimedia transmission in WMSNs. All proposed cross layer protocols for WMSNs have different objectives and different approach to reach these objectives, but still need more effort to satisfy QoS requirements for multimedia with the capability of WMSNs nodes.

Delay guarantee is very important especially in multimedia data, delay may occur in many layers, for that control the congestion at transport layer, select the best path routing at network, scheduling the packets at Mac layer and give high priority for important packets. Add cache to sensor nodes to store data temporary in case of lost packet. Residual nodes energy is important during discover the route to avoid path disconnect.

REFERENCES

- [1] I. F. Akyildiz and M. C. Vuran, *Wireless sensor networks* vol. 4: John Wiley & Sons, 2010.
- [2] H. Karl and A. Willig, *Protocols and architectures for wireless sensor networks*: John Wiley & Sons, 2007.
- [3] L. D. Mendes and J. J. Rodrigues, "A survey on cross-layer solutions for wireless sensor networks," *Journal of Network and Computer Applications*, vol. 34, pp. 523–534, 2011.
- [4] T. Melodia and I. F. Akyildiz, "Research Challenges for Wireless Multimedia Sensor Networks," in *Distributed video sensor networks*, ed: Springer Science & Business Media, 2011, pp. 233–246.
- [5] Z. Hamid and F. B. Hussain, "QoS in wireless multimedia sensor networks: a layered and cross-layered approach," *Wireless personal communications*, vol. 75, pp. 729–757, 2014.
- [6] Z. Hamid, F. Bashir, and J. Y. Pyun, "Cross-layer QoS routing protocol for multimedia communications in sensor networks," in *Ubiquitous and future networks (ICUFN), 2012 fourth international conference on*, 2012, pp. 498–502.
- [7] H. Wang, W. Wang, S. Wu, and K. Hua, "A survey on the cross-layer design for wireless multimedia

- sensor networks,” in *Mobile Wireless Middleware, Operating Systems, and Applications*, ed: Springer, 2010, pp. 474–486.
- [8] M. O. Farooq, M. St-Hilaire, and T. Kunz, “Cross-layer architecture for qos provisioning in wireless multimedia sensor networks,” *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 6, pp. 176–200, 2012.
- [9] M. AlNuaimi, F. Sallabi, and K. Shuaib, “A survey of wireless multimedia sensor networks challenges and solutions,” in *Innovations in Information Technology (IIT), 2011 International Conference on*, 2011, pp. 191–196.
- [10] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury, “Wireless multimedia sensor networks: Applications and testbeds,” *Proceedings of the IEEE*, vol. 96, pp. 1588–1605, 2008.
- [11] Z. Hamid and F. Bashir, “XL-WMSN: cross-layer quality of service protocol for wireless multimedia sensor networks,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, pp. 1–16, 2013.
- [12] S. Paniga, L. Borsani, A. Redondi, M. Tagliasacchi, and M. Cesana, “Experimental evaluation of a video streaming system for wireless multimedia sensor networks,” in *Ad Hoc Networking Workshop (Med-Hoc-Net), 2011 The 10th IFIP Annual Mediterranean*, 2011, pp. 165–170.
- [13] G. Sun, J. Qi, Z. Zang, and Q. Xu, “A Reliable Multipath Routing algorithm with related congestion control scheme in Wireless Multimedia Sensor Networks,” in *Computer Research and Development (ICCRD), 2011 3rd International Conference on*, 2011, pp. 229–233.
- [14] S.-Y. Bae, S.-K. Lee, and K.-W. Park, “Cross-layer QoS architecture with multipath routing in wireless multimedia sensor networks,” *International Journal of Smart Home*, vol. 7, pp. 219–226, 2013.
- [15] T. Çevik and A. H. Zaim, “A multichannel cross-layer architecture for multimedia sensor networks,” *International Journal of Distributed Sensor Networks*, vol. 2013, 2013.
- [16] G. H. E. Fard, M. Yaghmaee, and R. Monsefi, “An adaptive cross-layer multichannel QoS-MAC protocol for cluster based wireless multimedia sensor networks,” in *Ultra Modern Telecommunications & Workshops, 2009. ICUMT’09. International Conference on*, 2009, pp. 1–6.
- [17] G. EkbataniFard, M. H. Yaghmaee, and R. Monsefi, “A QoS-Based Multichannel MAC Protocol for Two-Tiered Wireless Multimedia Sensor Networks,” *Int’l J. of Communications, Network and System Sciences*, vol. 3, p. 625, 2010.
- [18] L. Shu, Y. Zhang, Z. Yu, L. T. Yang, M. Hauswirth, and N. Xiong, “Context-aware cross-layer optimized video streaming in wireless multimedia sensor networks,” *The Journal of Supercomputing*, vol. 54, pp. 94–121, 2010.
- [19] G. Shah, W. Liang, and X. Shen, “Cross-layer design for QoS support in wireless multimedia sensor networks,” in *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*, 2010, pp. 1–5.
- [20] G. Shah, W. Liang, and O. B. Akan, “Cross-layer framework for QoS support in wireless multimedia sensor networks,” *Multimedia, IEEE Transactions on*, vol. 14, pp. 1442–1455, 2012.
- [21] D. Rosário, R. Costa, H. Paraense, K. Machado, E. Cerqueira, T. Braun, et al., “A hierarchical multi-hop multimedia routing protocol for wireless multimedia sensor networks,” *Network Protocols and Algorithms*, vol. 4, pp. 44–64, 2012.
- [22] D. Rosário, R. Costa, A. Santos, T. Braun, and E. Cerqueira, “QoE-aware Multiple Path Video Transmission for Wireless Multimedia Sensor Networks,” *Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos—SBRC*, pp. 31–44, 2013.
- [23] J. Guo, L. Sun, and R. Wang, “A Cross-layer and Multipath based Video Sensor Transmission Scheme for Wireless Multimedia Sensor Networks,” *Journal of Networks*, vol. 7, pp. 1334–1340, 2012.
- [24] J. Zhang and J. Ding, “Cross-layer optimization for video streaming over wireless multimedia sensor networks,” in *Computer Application and System Modeling (ICCASM), 2010 International Conference on*, 2010, pp. V4–295–V4–298.
- [25] I. T. Almalkawi, M. Guerrero Zapata, and J. N. Al-Karaki, “A cross-layer-based clustered multipath routing with QoS-aware scheduling for wireless multimedia sensor networks,” *International Journal of Distributed Sensor Networks*, vol. 2012, 2012.
- [26] Z. Hamid, F. B. Hussain, and J.-Y. Pyun, “Delay and link utilization aware routing protocol for wireless multimedia sensor networks,” *Multimedia Tools and Applications*, pp. 1–22, 2015.
- [27] A. K. Demir, H. E. Demiray, and S. Baydere, “QoS-MOS: cross-layer QoS architecture for wireless multimedia sensor networks,” *Wireless networks*, vol. 20, pp. 655–670, 2014.
- [28] Y. Ozen, C. Bayilmis, N. Bandirmali, and I. Erturk, “Two tiered service differentiation and data rate adjustment scheme for WMSNs cross layer MAC,” in *Electronics, Computer and Computation (ICECCO), 2014 11th International Conference on*, 2014, pp. 1–4.
- [29] L. You and C. Liu, “Robust cross-layer design of wireless multimedia sensor networks with correlation and uncertainty,” *Journal of Networks*, vol. 6, pp. 1009–1016, 2011.
- [30] N. M. C. Tiglao and A. M. Grilo, “Cross-layer caching based optimization for wireless multimedia sensor networks,” in *Wireless and mobile computing, networking and communications (WiMob), 2012 IEEE 8th international conference on*, 2012, pp. 697–704.
- [31] S. M. Aghdam, M. Khansari, H. R. Rabiee, and M. Salehi, “WCCP: A congestion control protocol for wireless multimedia communication in sensor networks,” *Ad Hoc Networks*, vol. 13, pp. 516–534, 2014.
- [32] S. Sridevi and M. Usha, “Towards a cross layer framework for improving the QoS of delay sensitive heterogeneous WMSNs,” in *Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on*, 2013, pp. 1–5.
- [33] P. Duan, L. Liu, and Z. Zhang, “A Cross Layer Video Transmission Scheme Combining Geographic Routing and Short-Length Luby Transform Codes,” *International Journal of Distributed Sensor Networks*, 2015.

OXLP: An optimized cross-layers protocol for wireless sensor networks

Ahlam Saud Althobaiti & Manal Abdullah

King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT: You Wireless Sensor Network (WSN) is used widely in modern networks to monitor physical and environmental situations, such as temperature, sound, etc. However, they are considered emerging technologies, WSNs used in a wide applications over open networks. WSN is built from thousands of nodes; each node is connected wirelessly to one or several sensors. On the other hand, the posed considered one of the major challenges in developing WSNs because of the limited resources of the sensor. However, the requirement of significant saving in energy consumption of the sensor nodes is considered the most challenge facing WSNs.

To achieve more energy efficient WSN, the cross-layered approach is proven to be effective more than in the traditional layered approach. Control overhead is highly reduced using cross-layered approach. The protocol stack is considered as a system not as individual layers which are independent of each other by sharing the information from the system in the cross-layered approach. The Medium Access Control (MAC) protocol in WSNs extremely affects the energy consumption for sensor nodes. There are many MAC protocols that have been successfully designed towards the prime objective of energy efficiency. However, the classical layered protocol approach is considered the vast majority of the approaches which the most of existing protocols are based on. This paper mainly contributes towards the design of cross-layer protocol, OXLP that joint optimal design of MAC and network layer tasks. This protocol considers, beside energy consumption, packet delay, packet delivery, traffic adaptability, scalability, etc., for sensor nodes. While OXLP protocol improves energy consumption over well-known protocols in same filed, also both of the packet delivery ratio and packet delay reach 777ed at good level compared with other protocols in literatures. It is proven its efficiencies for traffic adaptability and scalability.

1 INTRODUCTION

The computing and communication technologies have been widely developed specifically in the last decades. In 1972, Gordon Bell expected that every decade the world would have new generation of computing technology. This improvement during the current fifth generation in the integration scale has mostly earned everything for the computing and communication technologies; such as reducing the cost, shrinking the size, reducing the switching power consumption, increasing the speed and efficiency, and providing the mobility and portability features.

The wireless networks with sensors integration developed this technology domain by making data movement, network distance, and network monitoring seamless. Wireless networks are flexible by allowing users to get connection anywhere with no more restriction of cables cost. The invisibility feature in the embedded systems can be integrated into the environment by assisting users in performing their tasks. Therefore, the compatibility between these advancements has improved small devices to re-organize themselves, and also it has introduced the domain of Wireless Sensor Networks (WSNs).

A new perspective has been added to the wireless technologies depending on the pervasiveness and self-organization of low-cost, low-power, long-lived, and small-sized sensor node.

In addition WSNs are sensing, computation, and communication into a single tiny device, furthermore, for several used applications they are emerging as an ideal candidate. Particularly, this emerging is highlighting in monitoring and controlling domains. In general, the networks demands for improvement are exponentially expanding with the increase in the networks dimensions [1].

WSNs have their own design and resource constraints, they are unlike traditional networks. The design constraints are dependent on application and are based on monitored environment [2]. Whatever the design approach, it is essential that WSN are subject to a rigorous analysis to provide long-term survivability of the architecture. The OSI (Open Systems Interconnection) layer model is generally used to specify the protocol architecture. However, in WSNs, it becomes difficult to use the traditional layer model. Cross-layer design is proposed to achieve an enhancement in performance for overall system in wireless networks [3].

Cross-layer techniques proposed improve energy conservation in WSN [2]. They use MAC (Media Access Control) layer information such as joint scheduling, power control, and sleep state of sensor nodes, to control energy consumption. The common mechanism used is to turned the node transceiver into a low power sleep state when it is not being used.

This research mainly contributes towards the design of a novel cross-layer protocol OXLP (An Optimized Cross-Layers Protocol). Proposed protocol is characterized by mainly integrating the functionality of MAC layer and network layer with a view towards inclusion of higher layers as well. The proposed protocol includes features from both MAC and network layers whereas it significantly reduces energy consumption of nodes through increase the sleep periods as much as possible, dealing with collision and control overhead. At the same time, it substantially aims to maintain packet delay as low as possible by enabling the receiving node to respond early and adaptively to the sending node.

The performance of the proposed OXLP protocol was evaluated through simulations. Using MATLAB A [4], simulator experiments are designed and implemented. The effectiveness of OXLP protocol is demonstrated in terms of packet delivery ratio, network lifetime, delivery delay to the BS and consumed energy for various load of traffic in the sensors network.

The rest of this work is organized as Section 2 explains the background, Section 3 shows the proposed protocol by explain the OXLP protocol, Section 4 the simulation experiment design for evaluation the OXLP protocol and Section 5 concludes the paper by presenting conclusion and directions for future research.

2 RELATED WORK

As proposed in [5], [6], [7] and [8], these studies are basically focused on MAC layer. However, working with single layer may lead to inefficient utilization for network resources. Recently, the camping between cross-layer design approach and TDMA scheduling is to obtain prolonged network lifetime. Article [9] addresses three main titles which are: a cross-layer optimization problem of joint design of routing, Medium Access Control (MAC), and physical layer protocols with cooperative communication. As the aim of article [9] is majored to achieve a minimum power cost under specified per-hop Packet Error Rate (PER) in wireless sensor networks.

CL-MAC protocol [10] is a novel cross-layer MAC protocol. Significantly, it is different from other MAC protocols since it is supporting

construction of multi-hop flows. Moreover, all pending packets in the routing layer buffer and all flow setup requests from neighbours are considered in the CL-MAC, which will be occurred when setting up a flow in CL-MAC. These considerations allow CL-MAC to make more informed scheduling decisions, reflecting the current network status, and optimizing its scheduling mechanism dynamically.

Authors in [11] proposed a cross-layer optimized geographic node-disjoint multipath routing algorithm, that is, two phase geographic greedy forwarding plus. To optimize the system as a whole, their algorithm is designed on the basis of multiple layers' interactions, taking into account physical layer; sleep scheduling layer and routing layer.

In this context authors proposed cross layers protocol in [12], based on the combined use of a duty-cycling protocol and a new kind of active wake-up circuit, based on a very-low-consumption Radio Frequency (RF).

Although various MAC protocols have been proposed, there is a possible future work for system performance optimization such as; Cross-layer optimization, Cross-layer interaction, etc. Interaction with the MAC layer can provide other layers with congestion control information; as well it can enhance routing selection. Many existing MAC protocols have been successfully addressed to present the performance studies of the static sensor nodes, but still there is a lack of literature for comparing these protocols with mobile networks. However, enhancing the MAC protocol can significantly improve communication reliability and energy efficiency.

3 OXLP: AN OPTIMIZED CROSS-LAYERS PROTOCOL

Figures In WSN domain vast majority of proposed protocols are based on approach of one-layer model of stack. In recent times, some works tend to exploit together many layers in order to optimize the network performance. This section proposed An Optimized Cross-Layers Protocol (OXLP). It is based on (MAC layer and Network layer) that are two adjacent layers to ensuring the pet performance for sensor network.

3.1 System model

3.1.1 Assumptions

For the sake of clarity, we first present some assumptions that hold in OXLP protocol. These assumptions are necessary to ensure network integrity and consistency. The following assumptions are made about the sensor nodes and the network model:

- The wireless network model is based on one Base Station (BS) or sink node and large number of sensors. As the BS is generally connected to the main power source, it is not a restrictive power assumption.
- The BS has the power of transmission which it is high enough to reach all sensors in the network.
- In OXLP, the sensor network is grouped into different clusters. Each cluster in turn composed of one Cluster Head (CH) and cluster member nodes.
- The respective CH responsible for getting the sensed data from cluster member nodes, and forwards it to BS by using of a multi-hop forwarding if necessary.
- Linear model is assumed with the distance between nodes as variable s , as shown in Figure 1.
- Each sensor has a unique identifier that is appended to the information field in the packet to identify the source of data.

3.1.2 Energy model

This paper assumes simple model for the radio hardware energy consumption as shown in Figure 2, where the transmitter node consumes the energy to run the radio electronics and the power amplifier, and the receiver node consumes the energy to run the radio electronics.

From Figure 2, let k (bits) is the packet size, and E_{elec} (Joule/bit) represents the consumed energy for transmitting or receiving a k -bit of data. Let ϵ_{amp} (Joule/bit/m²) denotes the energy consumed by the power amplifier at the transmitter for achieving acceptable bit energy to noise power spectral at the

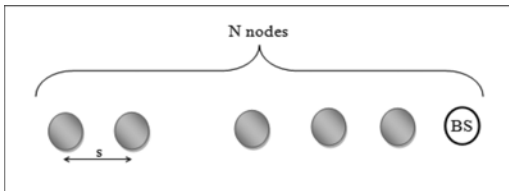


Figure 1. Simple linear sensor network.

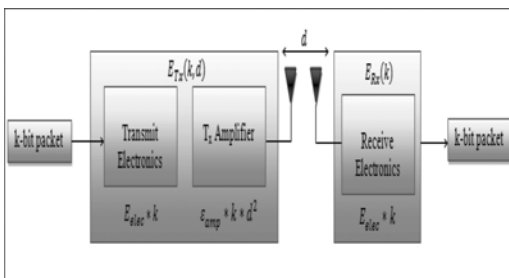


Figure 2. Radio energy consumption model.

receiver. Then, if source node x which is d far from its destination. transmits a k -bit packet, the radio dissipates as in the Equations 1 and 2:

$$E_{Tx}(k, d) = E_{Tx_elec}(k) + E_{Tx_amp}(k, d) \quad (1)$$

$$E_{Tx}(k, d) = E_{elec} * k + \epsilon_{amp} * k * d^2 \quad (2)$$

And to receive k -bit packet, the radio consumes energy as given by equations 3 and 4

$$E_{Rx}(k) = E_{Rx_elec}(k) \quad (3)$$

$$E_{Rx}(k) = kE_{elec} \quad (4)$$

The energy consumed by the radio during each idle listening period is expressed as equation 5

$$E_l(k) = \alpha E_{Rx}(k) \quad (5)$$

where α is the ratio of the energy consumed in receiving mode to the energy consumed in idle listening mode.

3.2 Overview

Based on joint functionalities of different underlying layers, the OXLP protocol is a Cross-Layer protocol which allows integrating MAC protocol and routing protocol for energy efficient delivery of data. The network layer uses information of data link layer when the routes establishment to access the medium efficiently, as shown in Figure 3. The forwarding process is composed of two phases: the MAC window and transmission window.

The distance between the respective cluster head and cluster member nodes is increased enormously

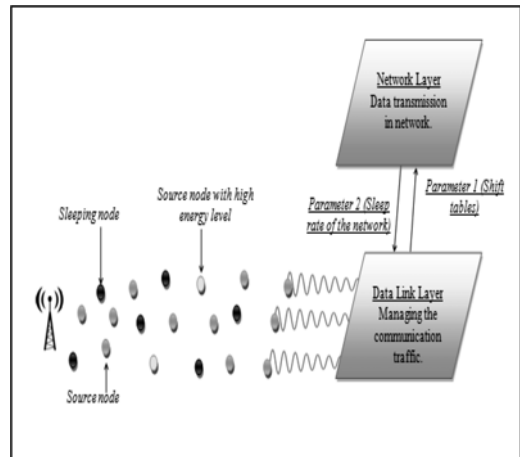


Figure 3. The Cross-layer optimized framework.

when the diameter of sensors network is increased beyond certain level. This case consumes a large transmission power of the node. This quickly consumes the nodes battery and reduces the system lifetime. To address the problem in this scenario, this paper propose multi-hop routing for OXLP protocol. Thus the sensors act as routers for other nodes' data in addition to sensing the data of environment.

There are two types of communication operations in OXLP protocol: Inter-cluster communication and intra-cluster communication. CH receives data from nodes and transmits this data to BS by using multi-hop if necessary in inter-cluster mode. On the other hand, in intra-cluster communication the nodes send their data directly to CH or through intermediate nodes during their allocated transmission slot.

Consequently total energy consumed in the sensor network might actually be less using OXLP protocol than direct transmission. To give more clarification, consider the linear sensor network as shown in Figure 1 since the average of distance between nodes is s . Consider energy consumed for transmitting a single k -bit message by using the direct communication approach from a node located a distance hs from the base station. From Equations 2 and 4, we have:

$$E_{Tx}(k, d = hs) = E_{elec} * k + \epsilon_{amp} * k * (hs)^2 \quad (6)$$

$$= k(E_{elec} + \epsilon_{amp} h^2 s^2) \quad (7)$$

where h is hops number and s is average of distance between the nodes.

In OXLP protocol, each node sends a message to the other node on the way to its CH. Also, CHs send to other CHs on the way to base station. Therefore, the nodes or CHs located a distance hs from its destination would require h transmits a distance s and $h-1$ receives.

$$E_{Tx}(k, d = hs) = hE_{Tx}(k, d = s) + (h-1) * E_{Rx}(k) \quad (8)$$

$$= h(E_{elec} * k + \epsilon_{amp} * k * s^2) + (h-1)kE_{elec} \quad (9)$$

$$= k((2h-1)E_{elec} + \epsilon_{amp} * hs^2) \quad (10)$$

where h is hops number and s is average of distance between the nodes.

3.3 OXLP protocol

The functional architecture of the OXLP protocol and the basic operation is illustrated in Figure 4. The operation of OXLP is divided into rounds. The MAC window is the beginning of the round where it organized the clusters and determined the routing paths. This is followed by transition window

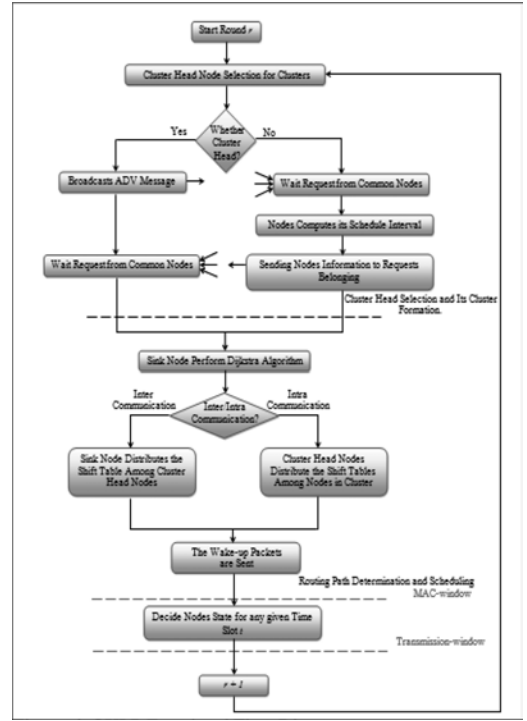


Figure 4. OXLP functional flow diagram.

where it transferred the data from the nodes to the cluster head then to the BS.

3.3.1 MAC window

MAC window introduces the core of OXLP protocol. The basic idea behind MAC window is to integrate both MAC and routing mechanism. This solution allows planning proactive routing table and medium access simultaneously. The routing table will be maintained by each cluster head, in which each entry contains destination ID, sender ID and allocated time slot.

This way affords three strong principles that are:

1. Allocate the time slots in efficient manner in order to avoid data collision. Simultaneously, it is fairly and efficiently shares the bandwidth of resources among multiple nodes of sensor for entire network.
2. In terms of network lifetime, the route of each message intended to the base station is selected by really crucial way.
3. Focus on increasing the sleep periods as much as possible, ensuring efficient awakening and avoid hidden and exposed terminal problems as proposed in [13].

In-depth detail of MAC window, it has two phases as follows:

1. Cluster formation and Cluster head selection.
2. Routing path determination and scheduling.

In next sub-sections, we will address MAC window phase in detail.

3.3.1.1 Cluster head selection and its cluster formation.

In OXLP protocol, CH selection phase apply same mechanism that is used in CSP sub-protocol in admin nodes selection sub-section, which introduced in detail in paper [13].

When clusters are created, each sensor node determines if it is become a cluster-head or not for the current round r . This decision is based on the suggested percentage of cluster heads for the sensor network which is determined a priori and also based on how many times the node has been a cluster-head so far.

If nodes elected as cluster-head for the current round, it must be broadcast an advertisement message to the rest of the nodes in its cluster. All other nodes other than the CH keep their receiver on and decide its CH. Every node selects a cluster head which is close to it. All nodes send its information to their respective CH. The CH creates a proper schedule for all sensor nodes in its cluster. Only during their respective schedules, nodes interact with neighbor nodes and the CH, else the nodes go to sleep mode. The cluster heads obtain sensed data from all nodes in its cluster, further aggregates the data and finally send it to the BS.

3.3.1.2 Routing path determination and scheduling

In this phase the routing path determination for intra-cluster as shown in Figure 5 and inter-cluster shown in Figure 6 communications. The determination of shortest path from sensor node to corresponding CH and from CH to BS is responsibility of MAC window, using Dijkstra's algorithm [14].

Each CH establishes and maintains traffic-based schedule information required by the transmitter and receiver selection for intra-cluster communications and BS responsible for scheduling the inter-cluster communications between all CHs in the network.

This schedule is defined as shift table in same way that in Schedule Protocol (SP) in MAC model which is introduced early.

These tables can be used to determine appropriate transmission, destination and sleep schedules for all sensor nodes, the transformation from source nodes to BS can be done efficiently in a collision free manner by such these information. To eliminating the need for a routing protocol, the shift tables themselves then serve to inherently form the routes through the sensor network.

3.3.2 Transmission window

In transmission window phase, CH collects data from all sensor nodes in its cluster and directly transmits the data, or through other cluster-head to BS.

Possible modes for each sensor node in transmission window are: Transmit (T_x), Receive (R_x), and Sleep (SL). To clarify, at any given slot t a node A is in the T_x mode if achieved one or both of the following conditions: (1) A has the highest priority to send its packet i.e., $prio(A, t)$ among its competing set and (2) A has packet to send. Regarding the R_x mode, it is achieved when node A is the intentional destination of the current transmitter. Otherwise, a node is switched off to the sleep SL mode. Therefore each node executes AP to decide its current mode (T_x , R_x , or SL) based on current node priorities and also on announced schedules by MAC window.

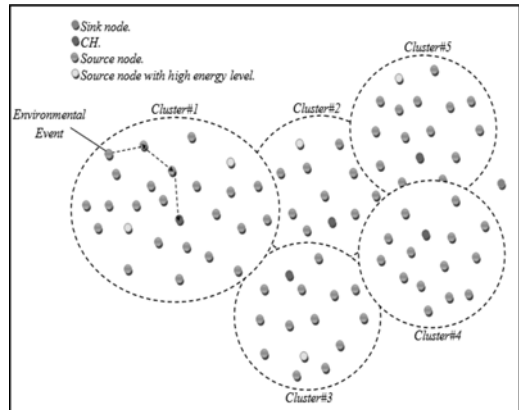


Figure 5. The intra-cluster routing.

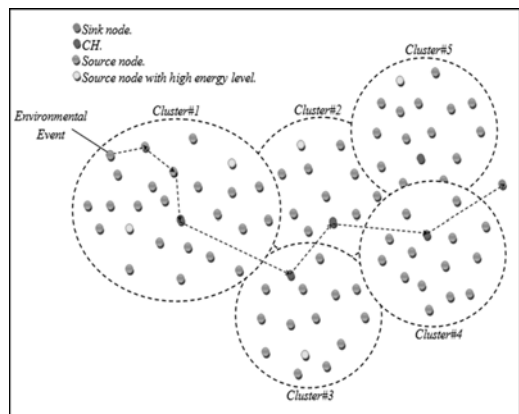


Figure 6. The inter-cluster routing.

The total power consumption at sensor node x , denoted by $E_{overall}$, is shown by Equation 11

$$E_{overall} = \sum_{i=1}^n E_{Tx(i)}(k, d) + \sum_{i=1}^n E_{Rx(i)}(k) + \sum_{i=1}^n E_{w(i)}(k) + \sum_{i=1}^n E_{ACK(i)}(k) + \sum_{i=1}^n E_{I(i)}(k) \quad (11)$$

where n : number of the scheduled packets, $E_{Tx}(k, d)$ is the energy consumed when source node x which is d far from its destination transmits a k -bit for n packet, $E_{Rx}(k)$ is the energy consumed to receive a n packet that has k -bit, $E_w(k)$ is the energy consumed in walk-up packets, $E_{ACK}(k)$ is the energy consumed for wait to receive the ACK value for data packets, and $E_I(k)$ is energy consumed by the radio during each idle listening mode.

3.4 Cross-layer optimization model

Based on information which is provided in section 3.3 a cross-layer optimization model can be formulated as the object function of Equation 12 as follows:

$$\begin{aligned} & \text{minimize}(E_{overall}) \\ & = \text{minimize} \left(\sum_{i=1}^n E_{Tx(i)}(k, d) + \sum_{i=1}^n E_{Rx(i)}(k) + \sum_{i=1}^n E_{w(i)}(k) + \sum_{i=1}^n E_{ACK(i)}(k) + \sum_{i=1}^n E_{I(i)}(k) \right) \end{aligned} \quad (12)$$

Now, as noted in proposed MAC protocol [13], $E_{I(i)}(k)$ is constant for a given WSN applications (i.e., given monitoring environment). For simplification, the sensor nodes in a cluster are in one of three possible states that are: transmit (TX), receive (RX), and sleep (SL). At any given time slot t , then the value of $E_{I(i)}(k) = 0$ because proposed MAC algorithm in OXLP protocol does not contain the idle state for any given t slots. Therefore we can consider the last summation term in Equation 12 as zero from these preceding statements. As a result for the foregoing, Equation 12 reduces to Equation 13 besides the fact that the beginning the energy optimization must done with respect to the backbone network nodes. Equation 13 as follow

$$E_{overall} = \sum_{i=1}^n E_{Tx(i)}(k, d) + \sum_{i=1}^n E_{Rx(i)}(k) + \sum_{i=1}^n E_{w(i)}(k) + \sum_{i=1}^n E_{ACK(i)}(k) \quad (13)$$

Furthermore, in case of best situation since the power consumption is reduced by decreasing over heading problem. This situation is also known as wining case as the communication can win time slots for data transmissions instead of sending walk-up packets. Based on what stated in previous, $E_{w(i)}(k)$ in Equation 13 is reduced and then this can leads to achieve the optimization model for OXLP protocol.

3.4.1 Proclamation 1

The energy consumption model proposed in (12) can be optimal on any link $(x, y) \in L(L$ is a set of links between nodes), the model must use the optimal transmission power of this link $E_{opt_Tx}(k, d)$ to achieve networkwide optimal energy consumption.

Proof. Assume that the network-wide per k -bit optimal relay energy consumption is

$$E_{opt_total}(k, d) = X + E_{Tx}(k, d) \quad (14)$$

where $E_{Tx}(k, d)$ is the per k -bit transmit energy dissipated on link x, y and X is the per k -bit transmit energy dissipated by the other links in the sensor network.

Suppose the transmission power at link (x, y) , $E_{Tx}(k, d)$, is not equal to the optimal transmission power $E_{opt_Tx}(k, d)$ of this link, we then have

$$E_{Tx}(k, d) > E_{opt_Tx}(k, d) \quad (15)$$

Thus,

$$\begin{aligned} & (E_{opt_total}(k, d) = X + E_{Tx}(k, d)) > \\ & (E''_{opt_total}(k, d) = X + E_{opt_Tx}(k, d)) \end{aligned} \quad (16)$$

Which contradicts with the statement that $E_{opt_total}(k, d)$ is the network-wide per k -bit optimal transmit energy consumption.

4 SIMULATION-BASED PERFORMANCE EVALUATION

In this section, the performance of the proposed approach is evaluated through simulation. A simulation is designed and implemented in MATLAB [4] to facilitate investigating the efficiency of OXLP protocol. In 4.2 sub-section, the research evaluates the performance of OXLP protocol and compares it against both cross-layer based protocols which are found in the literature such EYES [15] and PLOSA [16] and also against routing protocols for instance an application-specific protocol architecture for wireless micro-sensor networks LEACH [17].

4.1 Performance matrixes and simulations parameters

The proposed OXLP protocol is analyzed in terms of packet delivery ratio, network lifetime, delivery delay to the BS and consumed energy mechanism for various traffic loads also in term of control packet ratio. The load is expressed as the average number of new packets per slot. It can be easily expressed as a function of λ , an inter-arrival period of messages for a node. In case of the highest rate of 1-s inter-arrival time; the wireless channel is nearly fully utilized.

The paper assumes the same energy consumption is needed to send k-bits from A to B and vice versa. Table 1 summarize the parameters that used in the MATLAB simulator.

4.2 OXLP performance evaluation

4.2.1 Optimum number of clusters

Before proceeding for OXLP analysis it is important to determine the optimum number of clusters in the WSN. The number of cluster-heads in the network is key factor influencing the performance of network. Therefore, it's worth to do research on cluster head aspect.

As in LEACH [17], the application-specific protocol architecture for wireless micro-sensor networks LEACH designers observe the optimal value for clusters number per round to achieve the best performance. In LEACH protocol [17], the optimum number of clusters p_{opt} for a cluster-based sensor network has been illustrated in Equation 17, where N is the number of nodes that distributed uniformly in an $M \times M$ region.

$$p_{opt} = \frac{\sqrt{N}}{\sqrt{2\pi}} \sqrt{\frac{\epsilon_{fs}}{\epsilon_{mp}} \frac{M}{d_{to\ BS}^2}} \quad (17)$$

Table 1. The Simulation Parameters.

Parameter	Value
Number of sensor nodes	$n = 100$.
Packet size	$k = 4000$ bits
Network Area	$A = M \times M = 100 \times 100$
GW-node Location	Center Sink (50,50) Corner Sink (10,10)
Communication model	Bi-direction
Transmitter/Receiver Electronics	$E_{elec} = 50$ nJ/bit
Initial energy for normal node	$E_o = 0.5$ J
Data aggregation energy	$E_{DA} = 5$ nJ/bit/message
Transmit amplifier	$\epsilon_{amp} = 10$ pJ/bit/m ²

where ϵ_{fs} and ϵ_{mp} are power amplifier (free space (ϵ_{fs}) model and multipath (ϵ_{mp}) model).

In Equation 17, $d_{to\ BS}$ is the distance from the cluster-head node to the BS. The minimum and maximum values of $d_{to\ BS}$ is substituted, the upper and lower bounds of the desired number of clusters can be also obtained [17]. Though, p_{opt} will be selected regarding to:

- Average energy waste per round.
- Number of data packets which is received by BS per unit time that locates the network quality.

As in LEACH performance analysis, it takes 5% of the total number of nodes in WSN as the optimal number, as well the routing protocols also takes 5% for ideal working setting. In general, sensor nodes might be distributed in a large area, and some clusters might be not close to the BS, while others are closed. This case shows the great transmission energy waste when the nodes use to transmit data to BS.

This study obtains optimal number of clusters by simulation experiments. In this study, parameters as illustrated in Table 1 are used in simulation. Then by using different values for the percentage of nodes representing cluster heads, the total system energy consumption for the specific percentage has been obtained.

The total system energy consumption has been shown by Figure 7. Meanwhile, the total is presented as a function of the percentage of cluster heads for 100 nodes. As shown in the graphs, the ideal percentage of nodes that need to be cluster heads in order to get the minimum energy consumption is not exactly 5%, but it is around 3% - 5% for the given density. As the obtained results, the node percentage changes with the changing node density.

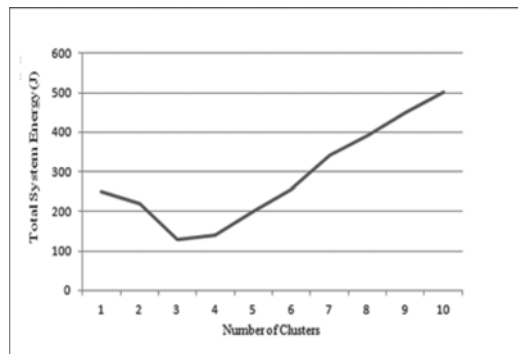


Figure 7. Energy dissipated OXLP protocol as the number of clusters is varied between 1 and 10. This graph shows that OXLP is most energy efficient when these are between 3 and 5 clusters in the 100-node network.

In case, there is only one cluster, the non-cluster head nodes often have to transmit data via long distance to reach the cluster head node, and this is draining their energy. As well as, if there are more than five clusters, then no much local data aggregation is being performed. However, for the rest of the research experiments, p_{opt} is set to 3%.

4.2.2 OXLP simulation results

The results of simulation show that OXLP protocol outperforms the MAC algorithm because of the improvement happening in the network layer in case of OXLP protocol. Figure 8(a) shows the average packet delivery ratio for OXLP protocol against proposed MAC algorithm and EYES protocol, PLOSA protocol and LEACH protocol in Figure 8(b).

Figure 9 shows the network life time for OXLP protocol against proposed MAC algorithm and other cross-layer based protocols.

In Figure 9(b), it is observed that the lifetime of network for OXLP protocol is the longest network lifetime while the network lifetime of LEACH protocol is the shortest ones.

Figure 10 shows end-to-end delay to the BS for OXLP protocol against proposed MAC algorithm and other cross-layer based protocols respectively.

LEACH protocol has higher delay. This is because the process of route discovery and queue in the data packet transmission. This causes a limitation of LEACH protocol.

Figure 11 shows the energy consumed for OXLP protocol against proposed MAC algorithm and other cross-layer based protocols respectively.

From the resulting routing scheme, Figure 11(b) shows that there obviously exist some redundant time slot allocations in EYES protocol and PLOSA protocol, which cause more energy consumption than necessary.

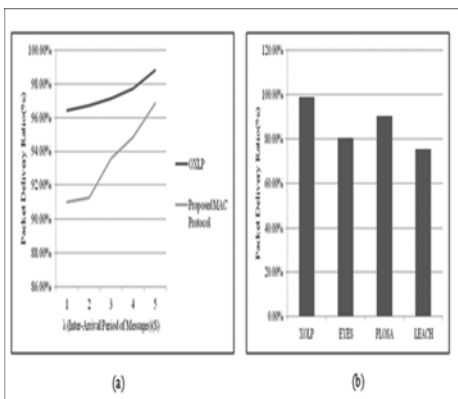


Figure 8. Packets delivery ratio. (a) Delivery ratio for proposed MAC protocol and XOLP protocol. (b) Delivery ratio for WSN protocols.

This is because the routing scheme for these protocols is functionality-oriented routing algorithm and the performance of these routing algorithms ignores energy consumption at nodes or in information transmission in a WSN. This means that, if the consumption of energy at a sensor node is high to an extent and the consumption of energy in any of sensor nodes is the same the energy transmission process can be ignored, the path that has the least number of sensor nodes from source node to destination node consumes the least energy. To overcome this drawback, we add shortest path routing scheme by Dijkstra algorithm [14] in OXLP protocol which in turn also caused more energy consumption in OXLP protocol.

4.2.3 OXLP scalability

Scalability is a significant factor in this study and should be highlighted. According to the network

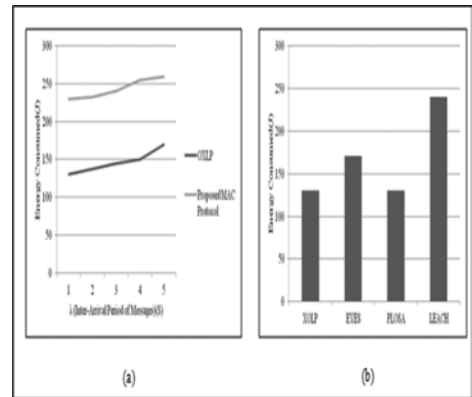


Figure 9. The network lifetime. (a) The network lifetime for proposed MAC protocol and XOLP protocol. (b) The network lifetime for WSN protocols.

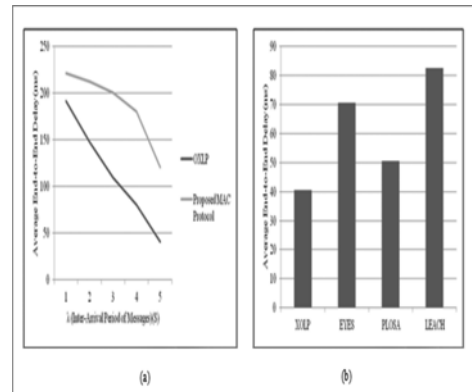


Figure 10. The Average end to end delay. (a) The average end to end delay for proposed MAC protocol and XOLP protocol. (b) The average end to end delay for WSN protocols.

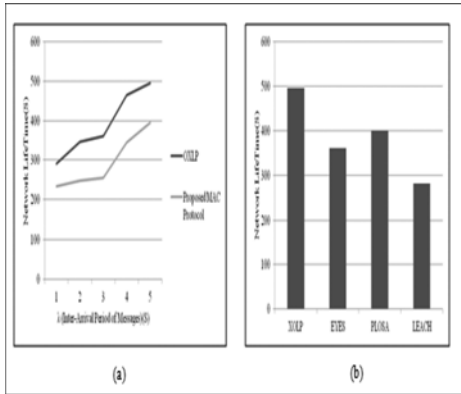


Figure 11. The energy consumed. (a) The energy consumed for proposed MAC protocol and XOLP protocol. (b) The energy consumed for WSN protocols.

growth or the workload, scalable protocol develops itself to suit the changes in the network size. Mainly, experiments focused on the node density that is based on different performance metrics. To analyze the performance of the OXLP protocol in scalability factor, some of the performance metrics in Table 1 are used. In the WSNs, more nodes should be alive to have high network lifetime, since, the results are really monitored based on parameters performance. However, protocol performance index is presented as network lifetime for analyzing OXLP.

– *Alive Node Vs Network Lifetime*

In fact, WSN demonstrates that the network application has been impacted by the active and monitor nodes. In addition, wireless sensor network have a limitation in its battery-power, knowing, the node reach to a status called dead node when its power level becomes less than threshold or equal to zero. Figure 12, presents the simulation results for network lifetime the First Node Dies (FND) vs. a live node, also from Figure 12, it can be shown that the network lifetime will be decreased when the node density increased. Meanwhile, if the node density is decreased from 1000 nodes to 100 nodes, then the lifetime of network will be increased. Thus, the density of sensor node should be always small to get best network lifetime.

Actually, the OXLP protocol has disadvantage, since each node maintains a route structure to each different destination address. As well as, OXLP protocol uses a lot of memory space, which hinders the efficiency in large size network. It is clear that, with high density network (1000 nodes), the network lifetime quickly reach zero. While with low density network (100–200 nodes) it takes long time for the network to die.

– *Data Vs Energy*

As shown by Figure 13, the relation between the node density and BS, whereas the increase of the node density will lead to increase the data received by BS.

Moreover, the network which has a minimum number of nodes actually dissipates less consumption of energy with an acceptable amount of data that can be received by BS. Since, Figure 13 shows when the network has 1000 nodes, it is consumed more energy with maximum amount of the data, while when there are 100 nodes in the network it is consumed less energy with minimum amount of data which is received by the BS among the considered configuration. Besides, in WSN, the OXLP protocol is a preferable choice in case of increasing the dense network.

4.3 *Comparison of WSNs protocols*

Comparison results between the proposed cross-layer approach OXLP and some other protocols, shows that EYES and PLOSA protocols have been optimized and perform low power consumption to ensure a node lifetime of several years on a single battery compared to the traditional approaches. In a dynamic network topology, a network lifetime of

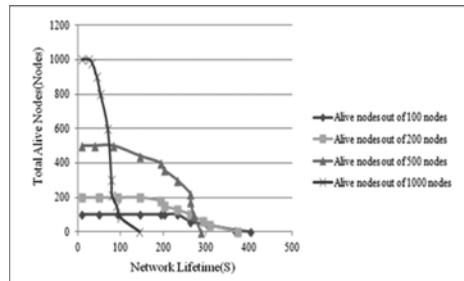


Figure 12. Alive Nodes Vs Network lifetime for different node density.

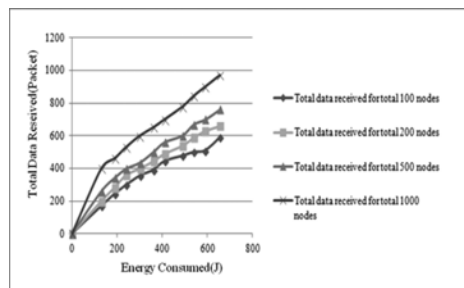


Figure 13. Data Vs Energy for different node density.

EYES has at least three times the lifetime of SMAC network. EYES performs better in scenarios where the nodes are mobile than in static cases. This can be explained by the fact that the roles active and passive are not changed in the latter case, while in the mobile case the dynamic changes in network topology force the nodes to reconsider their role. This leads to better and more even energy consumption between the nodes, which results in longer network lifetime. Since this protocol has a small standard amount of data reserved for route updates; in the static case this space is wasted. On the other hand, PLOSA distributes the node access in the frame according to their distance to the collector for multi-hop mechanism. The forwarding process is then simplified and can be done within a frame. Furthermore, PLOSA optimizes sleeping periods of devices because each node can receive packets to be forwarded only in a specific part of the frame. However, if two nodes send packets in parallel using PLOSA, one node delays its transmission and enters sleeping mode. Nodes stay longer in sleep mode than other modes. Whereby, micro-sensor network uses data aggregation locally to reduce the amount of transmitted data that reduces energy dissipation and latency in data transfer. Furthermore, adapting the clusters in micro-sensor approach depending on which nodes are cluster heads for a particular round (as in LEACH), this process is advantageous because it ensures that nodes communicate with the cluster head node that requires the lowest amount of transmit power. LEACH provides the high performance needed under the tight constraints of the wireless channel.

In OXLP protocol, the performance of the proposed cross-layer approach has also compared against cross-layer approaches. Hence, the proposed cross-layer approach improves energy conservation which performs high energy-efficient in WSN. OXLP protocol provides longer lifetime network. It uses an optimized MAC protocol that is based on TDMA and uses short-dynamic wake-up packets instead of the long preambles; these packets are carrying the ID for the intended node. Moreover, the proposed method assumes that all nodes sleep while the nodes is not scheduled to be active for sending or receiving data according to the presented shift table. Hence, shift table provide data routing table that enable the nodes in one cluster to be communicated based on its scheduled time slot without collision problem. OXLP protocol integrates both MAC and routing mechanisms to create an optimized routing table for data transmission in network clusters. However, the proposed OXLP protocol increases sleep stats, reduce overhearing, reduce overhead, and avoids collision problem. It determines shortest path routs from all sensor nodes to the corresponding CH in intra-cluster and between CH nodes to BS

node in the communication. Moreover, network changes should be handled rapidly and effectively for a successful adaptation: limited node lifetime and addition of new nodes to the network and varying interference which may alter the connectivity and then the network topology. Moreover, the proposed OXLP protocol performs high delivery rate for data with very low delay. The proposed approach may has some limitations to find shortest path in some cases of expanding network scalability, so the used shortest path algorithm may not apply to large networks' size as well as dynamic case due to its overwhelming additional works.

5 CONCLUSION AND FUTURE WORK

Several protocols have been presented in the literature review which is considered as ineffective in sensor networks based study.

The design and optimization of cross-layer considered as a new technique that can be applied to improve the performance of sensor networks. The main idea behind cross-layer design is to optimize the control and exchange of information over more than one layer. By exploiting the interactions between various protocol layers, this optimization, and leads to significant improvements of performance.

The Optimized Cross-Layers Protocol (OXLP) is developed in this research to provide an efficient communication method for wireless sensor networks. The proposed OXLP protocol based on the integration of MAC protocol and routing protocol for energy efficient data delivery to the sink node. Furthermore, the proposed protocol considers an optimization by involving the medium access control layer and network layer.

The simulation experiments showed the effectiveness of OXLP protocol in term of energy consumption (around 26.7% over EYES protocol and 59.4% over LEACH protocol). In respect of network lifetime, the OXLP protocol outperforms EYES protocol around 30.2% and PLOSA protocol around 21.2%. In term of end-to-end delay the OXLP protocol achieved improvement about 54.5% over EYES protocol and about 65.7% over LEACH protocol.

Other interesting characteristics of this protocol, that we can found in the proposed MAC method in this research that allow to mitigate some problems such as collision and idle listening which has been proved to be potential sources of energy wastage. Meanwhile the simulation experiment results show that OXLP significantly improved the communication performance and outperforms the proposed MAC protocol in terms of both network lifetime (around 22.5%) and consumed energy (around 41.3%) this improvement due to applying cross layering technique.

The overall conclusion is that OXLP protocol is best choice to move towards a network with less energy consumption as it involves energy minimizing techniques like multi-hop communication, clustering and data aggregation. Therefore, for applications where energy utilization is more critical like health monitoring, OXLP protocol is the best choice. OXLP protocol uses both inter cluster as well as intra cluster communication. For applications where network subjected to more scalability like environmental monitoring, OXLP protocol is good choice because it has the high delivery rate for data with low energy consumed and no matter how large the network. The scalability in OXLP protocol can further improved by improving the routing technique.

As a future work, the proposed protocol has to be formally validated; the experimentation on real sensors has to be performed in order to verify the performances of proposed protocol. Also, we will improve real-time property (delivery ratio) under harsh conditions. This can be an issue in highly critical applications. Two solutions can be explored to address this issue. First, work can be done at the physical layer in order to make the radio links more reliable by optimizing decoding thresholds for example (selection of the best links while keeping connectivity). Another solution could be an algorithm which reserves a good path in terms of links quality from the source to the sink, but this would imply more signalization thus more energy consumption.

REFERENCES

- [1] M. Yao, C. Lin, P. Zhang, Y. Tian, and S. Xu, "TDMA scheduling with maximum throughput and fair rate allocation in wireless sensor networks," in *2013 IEEE International Conference on Communications (ICC)*, 2013, pp. 1576–1581.
- [2] D. Espes, X. Lagrange, and L. Suárez, "A cross-layer MAC and routing protocol based on slotted aloha for wireless sensor networks," *Ann. Telecommun.*, pp. 1–11, Apr. 2014.
- [3] L. Shi and A. Fapojuwo, "TDMA Scheduling with Optimized Energy Efficiency and Minimum Delay in Clustered Wireless Sensor Networks," *IEEE Transactions on Mobile Computing*, vol. 9, no. 7, pp. 927–940, Jul. 2010.
- [4] "MATLAB - The Language of Technical Computing." [Online]. Available: <http://www.mathworks.com/products/matlab/>. [Accessed: 28-Dec-2014].
- [5] R. Ramaswami and K. K. Parhi, "Distributed scheduling of broadcasts in a radio network," in *Technology: Emerging or Converging, IEEE INFOCOM '89. Proceedings of the Eighth Annual Joint Conference of the IEEE Computer and Communications Societies*, 1989, pp. 497–504 vol. 2.
- [6] S.C. Ergen and P. Varaiya, "TDMA Scheduling Algorithms for Wireless Sensor Networks," *Wirel. Netw.*, vol. 16, no. 4, pp. 985–997, May 2010.
- [7] S. Chatterjea, L. F. W. Van Hoesel, and P. J. M. Havinga, "AI-LMAC: an adaptive, information-centric and lightweight MAC protocol for wireless sensor networks," in *Intelligent Sensors, Sensor Networks and Information Processing Conference, 2004. Proceedings of the 2004*, 2004, pp. 381–388.
- [8] L. Shi and A.O. Fapojuwo, "Cross-layer optimization with cooperative communication for minimum power cost in packet error rate constrained wireless sensor networks," *Ad Hoc Networks*, vol. 10, no. 7, pp. 1457–1468, Sep. 2012.
- [9] M.S. Hefaida, T. Canli, and A. Khokhar, "CL-MAC: A Cross-Layer MAC protocol for heterogeneous Wireless Sensor Networks," *Ad Hoc Networks*, vol. 11, no. 1, pp. 213–225, Jan. 2013.
- [10] D. Espes, X. Lagrange, and L. Suárez, "A cross-layer MAC and routing protocol based on slotted aloha for wireless sensor networks," *Ann. Telecommun.*, pp. 1–11, Apr. 2014.
- [11] G. Han, Y. Dong, H. Guo, L. Shu, and D. Wu, "Cross-layer optimized routing in wireless sensor networks with duty cycle and energy harvesting," *Wirel. Commun. Mob. Comput.*, p. n/a–n/a, Feb. 2014.
- [12] L. Catarinucci, R. Colella, G. Del Fiore, L. Mainetti, V. Mighali, L. Patrono, and M. L. Stefanizzi, "A Cross-Layer Approach to Minimize the Energy Consumption in Wireless Sensor Networks," *International Journal of Distributed Sensor Networks*, vol. 2014, p. e268284, Jan. 2014.
- [13] Ahlam Saud Althobaiti, Manal Abdullah, "Energy Efficient with Collision Free MAC Protocol for Wireless Sensor Network," *International Conference on Computer and Communication Engineering*, 2016.
- [14] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numer. Math.*, vol. 1, no. 1, pp. 269–271, Dec. 1959.
- [15] L. Van Hoesel, T. Nieberg, J. Wu, and P. J. M. Havinga, "Prolonging the lifetime of wireless sensor networks by cross-layer interaction," *IEEE Wireless Communications*, vol. 11, no. 6, pp. 78–86, Dec. 2004.
- [16] D. Espes, X. Lagrange, and L. Suárez, "A cross-layer MAC and routing protocol based on slotted aloha for wireless sensor networks," *Ann. Telecommun.*, pp. 1–11, Apr. 2014.
- [17] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 660–670, Oct. 2002.
- [18] A. Badi, I. Mahgoub, M. Slavik, and M. Ilyas, "Investigation of the effects of network density on the optimal number of clusters in hierarchical Wireless Sensor Networks (WSNs)," in *High-Capacity Optical Networks and Enabling Technologies (HONET), 2010*, 2010, pp. 171–177.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Classification for data stream clustering protocols in wireless sensor networks

Yassmeen Alghamdi & Manal Abdullah

Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT: The past few years have witnessed increased interest in the potential use of Wireless Sensor Networks (WSNs) in a wide range of applications and it has become a hot research area. Owing to the advances and growth in wireless communication technology, WSNs are becoming increasingly attractive for numerous application areas. Moreover, in some applications for sensor networks, the data usually arrives in an online fashion, is unlimited and there is no order in the arrival of data to be processed. Data with these characteristics are called data streams. This paper sheds the light on the most important concepts of WSNs, data stream mining, and data streams and clustering algorithms.

Keywords: WSNs, clustering, data streams, data mining

1 INTRODUCTION

For the last recent years, a widespread use of Wireless Sensor Networks (WSNs) have been seen in various applications. As known, a WSN is a special kind of the ad-hoc networks that have the ability to sense information and process them. They can be used in many fields such as environmental, industrial, military, and agriculture fields. Specifically, WSNs contain tiny independent built-in devices called sensor nodes. These sensor nodes contain four basic components: sensing unit, processing unit, transducer, and energy source [1]. Sensor nodes are mainly used in data processing and continuously report parameters such as temperature and humidity. Reports are transmitted by those sensors are collected by observers called Base Stations BS. WSNs depend hardly on their sensors that consumes a lot of battery. Unfortunately, the nature of WSNs make it very difficult to recharge the sensor node batteries. Therefore, energy efficiency is an important objective design in such networks [2]. WSN has several resource constraints, such as low computational power, limited energy source, and reduced bandwidth [1]. Therefore, WSNs algorithms should be accurately designed.

In some applications for sensor networks, data that WSNs process usually arrives in an online fashion. They are unlimited and there is no control on the arrival order of the elements being processed. Such data are called Data Streams [1, 3]. As a general rule, there are some differences between sensor streams and traditional streams. The sensor

streams are only samples of the entire population, imprecise, noisy, and with a moderate size. While in traditional streams the entire population is used, the data is exact, error-free, and huge [3].

The widespread deployment of WSNs and the need for data aggregation require efficient organization of the network topology to balance the load and extend the network lifetime. Clustering has proven to be an effective approach in solving the problem of energy consumption, data aggregation, scalability [4, 5] and organizing the network into a connected hierarchy. Generally, there are two categories of networks in WSNs, flat networks and hierarchical or clustered ones [6]. At any rate, clustering phenomenon plays an important role in the organization of networks, and also affects the network performance. Owing to a variety of advantages, clustering is becoming an active branch of routing technology in WSNs.

Due to the restrictions listed above, sending a large amount of data can take a lot of time when sensor nodes try to reach the wireless medium in a multi-hop data communication to get to the sink. For this reason, it is required to develop an algorithm for the generated data streams to reduce the network traffic that affects the data quality.

Trying to solve such limitations in sensor networks, it is the aim for researchers to propose distributed wireless sensor network data stream clustering algorithms to minimize sensor nodes energy consumption, extend the network lifetime and reduce data traffic to decrease the delay found in such networks. This paper briefly provides some

important concepts in WSNs, data streams, data stream mining, and clustering algorithms. Authors pay special attention to abovementioned concepts in WSNs.

The rest of this paper is organized as follows: section 2 represents an overview of data mining. Section 3 presents a survey on data clustering. Section 4 presents a discussion on data streams. Section 5 presents a classification for clustering protocols.

2 DATA MINING

Data mining is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics and database systems [7]. Large amount of databases in various areas have been generated from the development of information technology. The researches in databases and information technology have given the importance to store and manipulate with such data for further decision making. Data mining is a process of extracting information and patterns from huge data [8].

2.1 Common classes of data mining

There are six common classes of tasks for data mining [9]:

1. Anomaly detection unusual data records identification or data errors that require further investigation.
2. Association rule learning searches for relationships between variables.
3. Clustering is the task of discovering groups of data that are similar to each other.
4. Classification is the task of generalizing known structure to apply to new data.
5. Regression allows to find a function which models the data with the least error.
6. Summarization providing a compact representation of the data set, including visualization and report generation.

2.2 Data mining in WSNs

Today many organizations have a lot of large databases that grow without limit at a rate of several million records per day. Mining these continuous data streams brings new challenges [10].

Managing and processing data in WSNs has become a topic research in several fields of data mining. The main purpose of deploying the WSNs is to make the real-time decision which has been proved to be challenging due to many resource

constrained. This challenge helps the research community to find data mining techniques dealing with extracting knowledge from large continuous arriving data from WSNs. Traditional data mining techniques are not suitable for WSNs due to the nature of sensor data. [11].

2.2.1 Challenges of data mining in WSNs

Conventional data mining techniques for handling sensor data in WSNs are challenging for following reasons [11]:

1. Resource Constraint: The sensor nodes are resource constraints in terms of power, memory, communication bandwidth, and computational power.
2. Fast and Huge Data Arrival: The nature of WSNs data is its high speed. In many fields, data arrives faster than they could be mined. The challenge for data mining techniques is how to manipulate with the continuous, rapid, and changing data streams.
3. Online Mining: In WSNs, data is geographically distributed, inputs arrive continuously and so far newer data may change results based on older ones. Most data mining techniques that analyze data offline do not meet the requirement of handling distributed stream data.
4. Modeling Changes of Mining Results Over Time: Data-generating phenomenon is changing over time, so the extracted model should be updated continuously.
5. Data Transformation: Sensor nodes are limited in terms of bandwidth. So, transforming original data over the network is not easy.
6. Dynamic Network Topology: Sensor networks are deployed in harsh, uncertain, heterogenic, and dynamic environments. This can increase the complexity of designing an appropriate technique.

2.2.2 Taxonomy of data mining techniques for WSNs

There are three main classification levels in data mining techniques for WSNs [11]. The highest-level is based on general data mining classes used, such as frequent pattern mining, sequential pattern mining, clustering, and classification. For the clustering, it had adapted the K-mean, hierarchical, and data correlation-based. The second level of classification is based on the ability to process data on centralized or distributed manner. Since WSNs nodes has limited resource, the approach meant for distributed processing requires one-pass algorithms to complete a part of data mining locally, and then gather the results. The distributed approaches are used to increase the WSNs lifetime, and can extract a large number of data. The third

level [11] is selected based on how to face a specific problem. In WSNs, it has been focused on two aspects of issues: performance and application issues. Mainly, sensor nodes are constrained in some resources, so, algorithm that aware such constrains are needed to maximize the WSNs performance. On the other hand, a WSNs application requires data accuracy, fault tolerance, event prediction, scalability and robustness.

2.2.3 *Application areas of WSNs data mining*

The following are examples of real-world applications in WSNs data mining [11]:

1. In the environmental monitoring, sensors are deployed in an unattended region to monitor the natural environment.
2. For the health monitoring, patients are equipped with small sensors on multiple different positions of their body to monitor their health or behavior.
3. Sensors in object tracking are embedded in moving targets to track them in real-time.
4. WSNs are usually deployed in harsh environments. Sensor nodes are resource constrained especially in terms of power. Data mining techniques help to identify the faulty or dead nodes.
5. In data analysis, data mining techniques help to discover data patterns in a sensor network for a certain application.
6. In real-time monitoring, data mining techniques help to identify certain patterns and predict future events.

2.2.4 *Implementation of WSNs data mining*

Three main types are used for data mining implementation in WSNs [11]:

1. Evaluation Method. Analytical modeling, simulation and real deployment (unfeasible) are the most commonly used techniques to analyze the performance of data mining technique for WSNs.
2. Data Source. The dataset used to experimentally validate the proposed technique. Two types of datasets are used, synthetic and real.
3. Optimization Objective. WSNs are constrained in different resources. Techniques should consider those constraints but mostly they cannot efficiently cover all the performance metrics.

3 DATA CLUSTERING

Owing to the advances and growth in wireless communication technology, WSNs are becoming increasingly attractive for a lot of application areas. Thus, WSNs connects between the physical

world, the computing world and the human society. The clustering phenomenon plays an important role in affecting the network performance and organizing the networks as well. There are several key limitations in WSNs, that clustering schemes must consider [12]. A good clustering algorithm should be able to adapt to a variety of application requirements.

Data Clustering can be identified as grouping similar objects. In order to support data aggregation through efficient network organization, nodes can be partitioned into a number of small groups called clusters. Each cluster has coordinator called cluster head CH [2].

3.1 *Hierarchical clustering structure*

Grouping sensor nodes into clusters has been widely used to satisfy the scalability objective and achieve high energy efficiency to prolong network lifetime in large-scale WSNs. The hierarchical routing and data gathering protocols imply cluster-based organization of the sensor nodes in order that data fusion and aggregation are possible, that leads to a great energy savings. In the hierarchical network structure each cluster has a leader called the Cluster Head (CH) that performs special tasks (i.e. fusion and aggregation) and several common Sensor Nodes (SN) as members [13].

The cluster formation leads to a two-level hierarchy where the cluster heads form the higher level and the sensor nodes form the lower level. The sensor nodes periodically transmit their data to the corresponding cluster heads. The cluster heads aggregate the data (thus decreasing the total number of relayed packets) and transmit them to the Base Station (BS). This can be directly or through the intermediate communication with other cluster heads. Cluster heads spend a lot of energy more than other sensor nodes due to sending the aggregated data all the time to higher distances. A common solution to balance the energy consumption among all the nodes in the network, is to periodically re-elect new cluster heads in each cluster [13].

The BS is the point of data processing for the data received from the sensor nodes, and where the data is accessed by the end user.

3.2 *Designing clusters in WSNs*

There are several key attributes that designers must carefully consider when designing clustered wireless sensor networks [12]:

1. Cost of Clustering: Resources other than network organization should be considered, such as communication and processing.

2. Selection of CHs: Depending on certain application, some requirements may play an important role in its operations.
3. Real-Time Operation: Some applications such as habitat monitoring, simply receiving data is enough for analysis. Other applications like military tracking, the real-time data acquisition is much more vital.
4. Synchronization: Slotted transmission schemes, allow nodes to regularly schedule sleep intervals to minimize energy consumption.
5. Data Aggregation: In crowded networks, there are many nodes sensing similar data. Data aggregation allows distinguishing between sensed data and useful data.
6. Repair Mechanisms: WSNs are often vulnerable to node mobility, node death and interference that can result in link failure.
7. Quality of Service (QoS): Many QoS requirements in WSNs are application dependent. It is important to consider these metrics when choosing a clustering scheme.

3.3 Clustering parameters

Clustering has many parameters, listed as follows [13, 14]:

1. Number of clusters (cluster count): Cluster count is a critical parameter with regard to efficiency of total routing protocol.
2. Intra-cluster communication: The communication between a sensor and its CH is assumed to be one-hop communication. However, multi-hop communication is required when the communication range is limited.
3. Nodes and CH mobility: when CHs or nodes are assumed to be mobile, the cluster membership for each node should dynamically change. On the other hand, stationary CH tends to yield stable clusters and facilitate intra-cluster and inter-cluster network management.
4. Nodes types and roles: In heterogeneous networks, CHs are able to have more computation and communication resources. On the other

hand, in homogeneous networks, all nodes have same capabilities and some are designated as CHs.

5. Cluster formation methodology: Clustering mostly is performed in a distributed manner without coordination. In few earlier approaches, a centralized approach uses one or more coordinator nodes to partition whole network off-line.
6. Cluster-head selection: CHs can be preassigned in heterogeneous environments. In most cases, for homogeneous environments, CHs are selected from the deployed set of nodes.
7. Stability: Clustering scheme is said to be adaptive when the cluster count changes and the node's membership evolves overtime. Otherwise, it is considered fixed.
8. Multiple levels: The concept of a multi-level cluster hierarchy provides better energy distribution and total energy consumption.

Table 1, compares between some clustering algorithms.

4 DATA STREAMS

Sensor networks are the key to gather information needed by smart environments. In many emerging applications, huge data streams are monitored in a network environment. Each sensor generates a data stream where new data entries keep arriving in a continuous manner [15].

Data that WSNs process usually arrives in an online fashion, is unlimited and there is no control in the arrival order of the elements to be processed [1, 3]. However, there is a difference between sensor stream and traditional stream. Data streams arrive continuously, thus clustering algorithms have to perform in a single scan.

An important characteristic of data streams is mining them in a distributed fashion. Individual processors may have limited processing and memory. Examples of such cases include sensor networks, in which it may be desirable to perform

Table 1. Some clustering algorithms with comparing parameters.

Protocol	CH selection	Node mobility	Clustering methodology	Multiple levels
LCA	ID-based	Possible	Distributed	No
LEACH	Random	Limited	Distributed	No
TL-LEACH	Random	Limited	Distributed	Yes
GROUP	Proximity	No	Hybrid	No
EECS	Energy	No	Distributed	No
WCA	Weight-based	Yes	Distributed	No
ACE	Connectivity	Possible	Distributed	No
LEACHC	Random	Limited	Centralized	No

in network processing of data stream with limited processing and memory [16].

4.1 *Traditional data mining and data stream mining*

The traditional data mining is centralized, computationally expensive, and focus on disk-resident transactional data. It collects data at the central site. On the other hand, the WSNs data flows continuously in systems with varying update rates. It is impossible to store the entire WSNs data or to scan through it multiple times [11].

4.2 *Data stream characteristics*

Data stream has different characteristics of data collection to the traditional database model. Such characteristics are when the data stream arrives, it isn't easy to be controlled by the order. Another characteristic, date of data stream continuous generation with time progresses. Also, a data stream is dynamic. Additionally, data stream can be read and process based on the arrival order [17].

Based on that, the processing of the data stream requires first, each data element should be examined once at most. Second, each data element should be processed as soon as possible. Third, memory usage for mining data streams should be limited even though new data elements are continuously generated. Finally, results generated by online algorithms should be immediately available upon user request [17].

4.3 *Algorithms of data streams*

Due to the one-pass constraints on the data set, it is difficult to adapt arbitrary clustering algorithms to data streams. In the context of data streams, it may be better to determine clusters in specific user defined horizons rather than on the entire data set. The micro-clustering technique determines clusters over the entire data set [16].

To be more general, there are many algorithms found to handle the data streams through various environments. Such algorithms can be listed as: Data Stream Clustering, Data Stream Classification, Frequent Pattern Mining, Change Detection in Data Streams, Stream Cube Analysis of Multi-dimensional Streams, Load-shedding in Data Streams, Sliding Window Computations in Data Streams, Synopsis Construction in Data Streams, Join Processing in Data Streams, Indexing Data Streams, Dimensionality Reduction and Forecasting in Data Streams, Distributed Mining of Data Streams, and Stream Mining in Sensor Networks.

5 CLUSTERING PROTOCOLS

Clustering protocols could be generally classified into three main types in our research scope. First, clustering routing protocols in WSNs (without data streams). Second, clustering protocols for data streams. Third, clustered WSNs for data streams.

5.1 *Clustering routing protocols in WSNs*

Based on network structure, routing protocols in WSNs can be divided into two categories: Flat and hierarchical routing. In a flat network topology, all nodes perform the same tasks and have the same functionalities. Data transmission is performed hop by hop using flooding form. The typical flat routings in WSNs include Flooding and Gossiping, SPIN, Directed Diffusion (DD), Rumor, GPSR, Trajectory Based Forwarding (TBF), Energy-Aware Routing (EAR), Gradient-Based Routing (GBR), and SAR. Flat routing protocols are effective in small-scale networks [6].

On the other hand, in hierarchical topology, nodes perform different tasks and are organized into lots of clusters according to specific requirements or metrics. Generally, CHs have the highest energy in the clusters to perform data processing and information transmission, while nodes with low energy act as Member Nodes (MNs) and perform the task of information sensing [6].

Clustering routings protocols in WSNs include LEACH, HEED, Distributed Weight-based Energy-efficient Hierarchical Clustering protocol (DWEHC), Position-based Aggregator Node Election protocol (PANEL), Two-Level Hierarchy LEACH (TL-LEACH), Unequal Clustering Size (UCS), Energy Efficient Clustering Scheme (EECS), EEUC, ACE, BCDCP, PEGASIS, Threshold sensitive Energy Efficient sensor Network protocol (TEEN), APTEEN, Two-Tier Data Dissemination (TTDD), Concentric Clustering Scheme (CCS), HGMR, etc.

5.2 *Proactive and reactive clustering*

Proactive clustering algorithms are based on the assumption that the sensors always have data to send, for that reason, they should all be considered during the cluster formation. On the other hand, reactive algorithms take advantage of user queries for the sensed data or of specific triggering events that occur in the WSN. Namely, nodes may react immediately to sudden hard changes in the value of a sensed attribute [13]. Figure 1, shows a flow-chart for some examples of "proactive" and "reactive" clustering protocols.

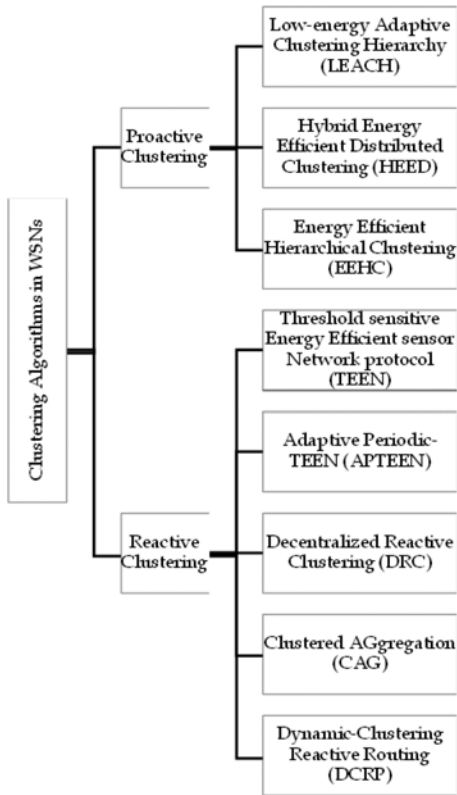


Figure 1. Examples of proactive and reactive clustering protocols.

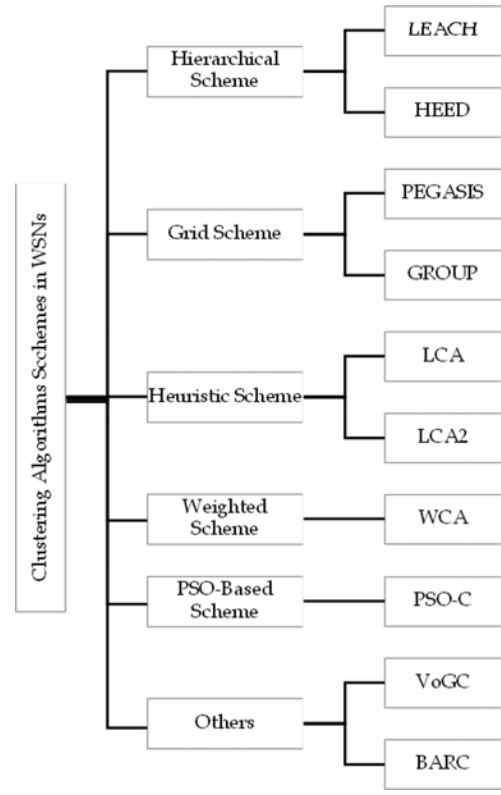


Figure 2. Clustering algorithms schemes in WSNs.

5.3 Clustering algorithms schemes in WSNs

Clustering algorithms could be considered under specific schemes, such schemes are hierarchical, grid, heuristic, weighted, PSO-Based and other schemes. Each scheme will be described in brief [12, 18, 19]. Figure 2 shows the clustering schemes in WSNs.

In LEACH, there are many rounds and each round has two main phases, a setup phase and steady state phase [19]. To reduce inter-cluster and intra-cluster collisions in LEACH, it uses a TDMA or CDMA MAC. The energy consumption of the information gathered by the sensors node will depend on the number of CHs and radio range of different algorithms [19]. Table 2 gives some examples of the LEACH descendants

Another protocol known as Hybrid Energy-Efficient Distributed Clustering (HEED) is a multi-hop clustering algorithm for WSNs. It focuses on efficient clustering by proper selection of CHs based on the physical distance between nodes. HEED has several objectives, where it distributes energy consumption to prolong network lifetime.

Table 2. Descendant of LEACH protocol.

Descendant of LEACH	Abbreviation
LEACH	Low energy adaptive clustering hierarchy
LEACH-C	Centralized-Low energy adaptive clustering hierarchy
LEACH-B	Balanced-Low energy adaptive clustering hierarchy
LEACH-ET	Energy threshold-Low energy adaptive clustering hierarchy
TL-LEACH	Three Layer-Low energy adaptive clustering hierarchy
Armor-LEACH	Advance LEACH routing protocol for micro-sensor networks
O-LEACH	Optical-Low energy adaptive clustering hierarchy
MR-LEACH	Multi-hop hop routing-Low energy adaptive clustering hierarchy
LEACH-D	Low energy adaptive clustering hierarchy-D

It also minimizes energy during the CH selection phase. Moreover, It can minimize the control overhead of the network. In HEED, CH selection is determined based on the residual energy [12, 18].

A Grid schemes

Power-Efficient GATHERing in Sensor Information Systems (PEGASIS) is a data-gathering algorithm that establishes the concept that energy savings can result from nodes not directly forming clusters. If nodes form a chain from source to sink, only one node in any given transmission time-frame will be transmitting to the base station. Data-fusion appears at every node in the sensor network allowing for all relevant information to permeate across the network. Moreover, the average transmission range required by a node to relay information can be much less than in LEACH [12, 18].

The Group algorithm is another grid-based clustering algorithm. In this algorithm one of the sinks, dynamically and randomly builds the cluster grid. Each new CH will then select more CHs along the grid until all CHs have been selected [12, 18].

B Heuristic algorithms

These algorithms has one or both of their goals during solving a problem. First, finding an algorithm with reasonable run-time. Second, finding the optimal solution. There are many types of heuristic algorithms that exist in choosing CHs.

Linked Cluster Algorithm (LCA) is one of the very first clustering algorithms developed. In LCA, each node is assigned to a unique ID and can become a CH if a node has the highest ID number or assuming none of its neighbors are cluster heads, then it becomes a CH [12, 18]. LCA2 was proposed to eliminate the election of an unnecessary number of CHs, as in LCA. LCA2 introduced the concept of a node being covered and non-covered. A node is covered when one of its neighbors is CH. CH election is done by starting with the node having the lowest ID among non-covered neighbors [12, 18].

C Weighted Schemes

Weighted Clustering Algorithm (WCA) is a non-periodic procedure to the CH election, invoked when every time a reconstruction of the networks topology is unavoidable. WCA tries to find a long-lasting architecture during first CH election. When a sensor loses the connection with its cluster head, the election procedure is invoked to find a new clustering topology. WCA is based on a combination of metrics such as: the ideal node degree, transmission power, mobility and the remaining energy of nodes [12, 18].

D PSO-Based scheme

In Centralized-PSO (PSO-C), nodes which have energy above average energy resource are elected as CHs. Simulation results show that PSO outperform

to LEACH and LEACH-C in terms of network life time and throughput [19].

E Other schemes

VoGC is a combination of voting method and clustering algorithm, developing new clustering schemes for secure localization of sensor networks. Voting-On-Grid Clustering (VOGC) is used instead of traditional clustering algorithms to reduce the computational cost. It is found that the scheme can provide good localization accuracy and identify a high degree of malicious beacon signals [19]. A mathematical battery model for implementation in WSNs was used in Battery Aware Reliable Clustering (BARC) algorithm. It improves the performance over other clustering algorithms due to using Z-MAC and rotating the CHs according to battery recovery schemes. Moreover, the BARC consists of two stages per round for selection of CH, initialization or setup and steady state [19].

5.4 Clustering protocols for data streams

In 2006, Feng Cao [20] proposed the DenStream algorithm for clustering dynamic data stream. It is an effective and efficient method that can discover clusters of arbitrary shape in data streams, but it is insensitive to noise [21]. The algorithm extends the micro cluster concept, and introduces the outlier and potential micro clusters to distinguish between real data and outliers.

Heng Zhu Wei [20] proposed a density and space clustering algorithm called CluStream. It is a data-stream clustering algorithm based on k-means that is inefficient to find clusters of arbitrary shapes and cannot handle outliers. Further, they require to know k and user-specified time window [22]. The DenStream and CluStream algorithms are not able to reveal clusters of arbitrary shape effectively and cannot distinguish clusters which have different levels of density [20].

K-means algorithm is used in the offline phase of some algorithms such as CluStream. It is a divide and conquer scheme that partition data streams into segments and discover clusters in data streams. The k-means has a number of limitations. First, it aims at identifying spherical clusters but is incapable of revealing clusters of arbitrary shapes. Second, it is unable to detect noise and outliers. Third, the algorithm requires multiple scans of the data, making it not directly applicable to large volume data stream [20, 22]. STREAM and CluStream are two well-known extensions of k-means on data streams [23]. Figure 3 shows some algorithms based on K-Means and Fuzzy C-Means.

Many recent data stream clustering algorithms are based on CluStream's two-phase framework. Wang et al. [26] proposed an improved offline component using an incomplete partitioning strategy.

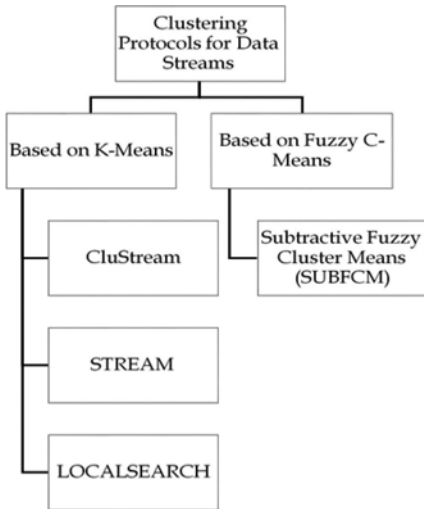


Figure 3. Clustering algorithms schemes in WSNs.

An extensions of this component including clustering multiple data streams, parallel data streams, distributed data streams and applications of data stream mining [22].

LOCALSEARCH, STREAM, DenStream and CluStream are clustering algorithms evolving data streams. They have ignored the problems of grid border. Data stream is coming with a large number in chronological order, and making the original grid no longer adapt to the new data mapping, so a large number of data is likely to fall on grid border. But if it is simply discarded, the cost will be greatly increased and the efficiency will be affected [20].

D-Stream is a density grid-based algorithm in which the data points are mapped to the corresponding grids and the grids are clustered based on the density [23].

MR-Stream is an algorithm that can cluster data streams at multiple resolutions. The algorithm partitions the data space into cells and a tree like data structure which keeps the space partitioning. The MR-Stream increases the performance of clustering by determining the exact time to generate the clusters [23].

FlockStream is density-based clustering algorithm based on a bio-inspired model. It uses the flocking model where agents are micro-clusters and they work independently. FlockStream merges online and offline phases. It can get the clustering results without performing offline clustering [23].

DenStream, MR-Stream, D-Stream and FlockStream are based on density-based clustering. They can affectively detect arbitrary shape clusters and handle noise, but their quality decrease when they are used for clusters with variant densities [23].

LOCALSEARCH algorithm uses dividing and conquering to partition data streams into segments, and discovers clustering of data streams in finite space, by using the k-means algorithm [20]. Later on, STREAM algorithm was proposed by O'Callaghan [23] which is based on LOCALSEARCH. It puts equal weights to outdated and recent data and cannot capture the evolving characteristics of data stream [20].

Data stream clustering analysis causes challenges for traditional clustering algorithms. The data can only be examined in one pass. Viewing data stream as a long vector of data is not enough in many applications [22].

Incremental DBSCAN is a method for data warehouse applications. It can only handle a relatively stable environment but it can't deal with limited memory and fast changing streams. HPSstream introduces the concept of projected cluster to data streams.[21].

A framework to dynamically cluster multiple evolving data streams called Clustering on Demand (COD) was proposed [28]. It produces a summary hierarchy of data statistics in the online phase, whereas the clustering is performed in the offline phase [24]. It summarizes the data streams using the Discrete Fourier Transform (DFT). An Online Divisive-Agglomerative Clustering (ODAC) was proposed to incrementally construct tree-like hierarchy.

Many density-based clustering algorithms are not suitable for data stream environments. They need two-pass of data and this condition is impossible for data streams. GMDBSCAN and ISDBSCAN use two-pass data. Other algorithms have high execution time which makes them not applicable for data streams. DSCLU is density-based clustering for data stream in multi density environments [23].

DD-Stream, is framework for density-based clustering stream data. The algorithm adopts a density decaying technique to capture the evolving data stream and extracts the boundary point of grid by using the DCQ-means algorithm. DD-Stream has better scalability in processing large-scale and high dimensional stream data [20].

D-Stream is a density-based clustering real-time stream data algorithm. It uses an online component which maps each input data record into a grid. It also has an offline component which computes the grid density [22].

5.5 Clustered WSNs for data streams

A distributed WSN data stream clustering algorithm called SUBFCM (Subtractive Fuzzy Cluster Means) was proposed to minimize sensor nodes energy consumption and extend the network lifetime. Simulations show that the energy efficient algorithm SUBFCM can achieve WSN data stream clustering with significantly less energy than that required by

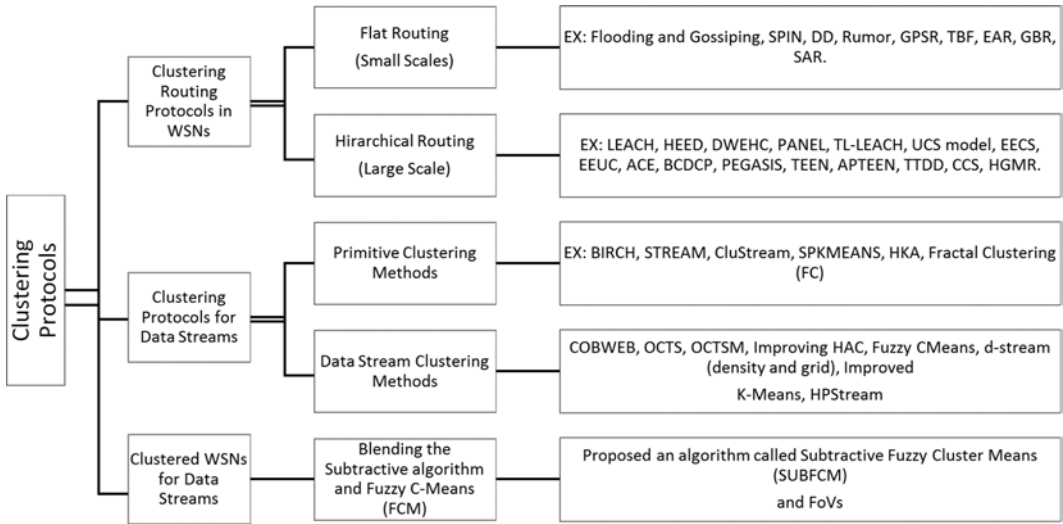


Figure 4. Classification of clustering algorithms.

known fuzzy c-means and k-means algorithms [25]. The proposed SUBFCM is a result of blending the Subtractive clustering and FCM algorithm. Fuzzy C-Means (FCM) is the most widely used algorithm in the field of data mining [25].

In Wireless Multimedia Sensor Networks (WMSNs), multimedia clustering protocols use the Quality of Service (QoS) parameters [26]. QoS has several metrics such as delay, bandwidth, reliability, jitter [27] and packet loss [26]. Many multimedia applications are time critical, they need to be reported with a limited time. The multimedia sensors have the ability to capture video, image, audio and scalar sensor data. A clustering algorithm for WMSNs has been proposed based on FoV areas. This algorithm aims to find the intersection polygon and computing the overlapped areas to establish clusters and determine cluster membership. For dense networks, overlapping FoVs causes consuming power of the system [19].

5.6 Taxonomy of clustering protocols

As mentioned previously, clustering protocols are classified in this research to three main types. To summarize, Figure 4 shows a structure that describes in brief the classification of clustering algorithms.

REFERENCES

- [1] A. L. de Aquino, C. M. S. Figueiredo, E. F. Nakamura, L. S. Buriol, A. Loureiro, A. O. Fernandes, *et al.*, "A sampling data stream algorithm for wireless sensor networks," in *Communications, 2007. ICC'07. IEEE International Conference on*, 2007, pp. 3207–3212.
- [2] O. Younis, M. Krunz, and S. Ramasubramanian, "Node clustering in wireless sensor networks: recent developments and deployment challenges," *Network, IEEE*, vol. 20, pp. 20–25, 2006.
- [3] A. L. de Aquino, C. M. Figueiredo, and E. F. Nakamura, "Data Stream Algorithms For Processing of Wireless Sensor Network Application Data."
- [4] O. Boyinbode, H. Le, A. Mbogho, M. Takizawa, and R. Poliah, "A Survey on Clustering Algorithms for Wireless Sensor Networks," in *2010 13th International Conference on Network-Based Information Systems*, 2010, pp. 358–364.
- [5] M. Abdullah, H. N. Eldin, T. Al-Moshadak, R. Alshaik, and I. Al-Anesi, "Density Grid-Based Clustering for Wireless Sensors Networks," *Procedia Computer Science*, vol. 65, pp. 35–47, 2015.
- [6] X. Liu, "A survey on clustering routing protocols in wireless sensor networks," *Sensors*, vol. 12, pp. 11113–11153, 2012.
- [7] S. Chakrabarti, M. Ester, U. Fayyad, J. Gehrke, J. Han, S. Morishita, *et al.*, "Data mining curriculum: A proposal (Version 1.0)," *Intensive Working Group of ACM SIGKDD Curriculum Committee*, 2006.
- [8] M. Bharati and M. Ramageri, "Data mining techniques and applications," 2010.
- [9] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI magazine*, vol. 17, p. 37, 1996.
- [10] P. Domingos and G. Hulten, "Mining High-Speed Data Streams," 2000.
- [11] A. Mahmood, K. Shi, S. Khattoon, and M. Xiao, "Data mining techniques for wireless sensor networks: A survey," *International Journal of Distributed Sensor Networks*, vol. 2013, 2013.
- [12] D. Dechene, A. El Jardali, M. Luccini, and A. Sauer, "A Survey of Clustering Algorithms for Wireless Sensor Networks."

- [13] B. Mamalis, D. Gavalas, C. Konstantopoulos, and G. Pantziou, "Clustering in wireless sensor networks," *RFID and Sensor Networks: Architectures, Protocols, Security and Integrations*, Y. Zhang, LT Yang, J. Chen, eds, pp. 324–353, 2009.
- [14] A. A. Abbasi and M. Younis, "A survey on clustering algorithms for wireless sensor networks," *Computer communications*, vol. 30, pp. 2826–2841, 2007.
- [15] E. Soroush, K. Wu, and J. Pei, "Fast and quality-guaranteed data streaming in resource-constrained sensor networks," in *Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing*, 2008, pp. 391–400.
- [16] C. C. Aggarwal, *Data streams: models and algorithms* vol. 31: Springer Science & Business Media, 2007.
- [17] L. Su, H.-y. Liu, and Z.-H. Song, "A new classification algorithm for data stream," *International Journal of Modern Education and Computer Science (IJMECS)*, vol. 3, p. 32, 2011.
- [18] R. Mitra and D. Nandy, "A survey on clustering techniques for wireless sensor network," *International Journal of Research in Computer Science*, vol. 2, p. 51, 2012.
- [19] V. Kumar, S. Jain, and S. Tiwari, "Energy efficient clustering algorithms in wireless sensor networks: A survey," 2011.
- [20] C. Jia, C. Tan, and A. Yong, "A grid and density-based clustering algorithm for processing data stream," in *Genetic and Evolutionary Computing, 2008. WGECC'08. Second International Conference on*, 2008, pp. 517–521.
- [21] F. Cao, M. Ester, W. Qian, and A. Zhou, "Density-Based Clustering over an Evolving Data Stream with Noise," in *SDM*, 2006, pp. 328–339.
- [22] Y. Chen and L. Tu, "Density-based clustering for real-time stream data," in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2007, pp. 133–142.
- [23] A. Amini, H. Saboohi, and T. Y. Wah, "A multi density-based clustering algorithm for data stream with noise," in *Data Mining Workshops (ICDMW), 2013 IEEE 13th International Conference on*, 2013, pp. 1105–1112.
- [24] J. Yin and M. M. Gaber, "Clustering distributed time series in sensor networks," in *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*, 2008, pp. 678–687.
- [25] H. Sabit, A. Al-Anbuky, and H. Gholam-Hosseini, "Distributed WSN data stream mining based on fuzzy clustering," in *Ubiquitous, Autonomic and Trusted Computing, 2009. UIC-ATC'09. Symposia and Workshops on*, 2009, pp. 395–400.
- [26] J. R. Diaz, J. Lloret, J. M. Jimenez, and J. J. Rodrigues, "A QoS-based wireless multimedia sensor cluster protocol," *International Journal of Distributed Sensor Networks*, vol. 2014, 2014.
- [27] M. Abazeed, N. Faisal, S. Zubair, and A. Ali, "Routing protocols for wireless multimedia sensor network: a survey," *Journal of Sensors*, vol. 2013, 2013.

An architecture for selling internet data using mobile hotspot

M. Taileb, B. Alshuaibi, W. Bagais, A. Basudan, N. Bahurmoz & M. Alsadi

Department of Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT: Nowadays, people try to efficiently utilize all their belongings, like renting their houses when they are in vacations, renting their cars when they don't use them or selling the goods they don't need. From this perspective came the idea of the proposed architecture; to sell your Internet data when you are not using it. People used to share their Internet data with others they know but the idea proposed in this paper is providing Internet access on the go, to people in need, in places with no public Internet access. In this paper, an architecture for selling the Internet data using the mobile Hotspot is proposed. The user's smartphone is used as an Internet access point using the smartphone Hotspot, this allows him to share his Internet connection with others but by selling them his Internet data. The proposed architecture has been implemented and successfully tested.

1 INTRODUCTION

People have Internet connection in their smartphones, at home and work. They pay for the Internet connections in their smartphones through communication companies. In many cases this results in an excess amount of internet data that is not used and wasted by the end of each month. On the other hand, there are people who might be in an urgent need to get connected to the Internet, e.g., to send or receive an important email. Those with excess (unused) Internet data can help people in need to Internet access by selling them their Internet data, via Wi-Fi Hotspot, with reasonable fees. In this case, both provider and consumer get advantages. The provider can sell his excess Internet data, and the consumer pays a small fee for a service he needs.

In literature, all research works are about Internet sharing and their related issues. In [1] and [2], sharing of home broadband connections with the public is discussed. In [3], authors proposed P2P architecture for large Wi-Fi sharing systems. In [4], authors investigate the users' attitudes towards Internet sharing. The main issue with Internet sharing is security, to address this issue; the main target of authors in [5], [6] and [7] is the strengthening of security. In [8], cooperative Wi-Fi-sharing networks are analyzed; where users cooperatively share their resources, such as wireless access points and Internet uplinks. An interesting work is proposed in [9], authors present a Wi-Fi sharing architecture, called Social WiFi, which enables WiFi sharing with online friends from online social networks.

However, in this paper, we propose a mobile application architecture where the user can use his smartphone as an Internet access point using the smartphone Hotspot, share his mobile Internet connection and sell his Internet data.

The rest of the paper is organized as follows. In section 2, the application architecture is described with its different stages. And in section 3, the experimental results are provided. Finally the section 4 concludes the paper.

2 APPLICATION ARCHITECTURE

A mobile application architecture using Wi-Fi Hotspot technology, called HotNet, is presented; it allows mobile users to sell their Internet data in places with no public Internet access. An overview of the application architecture is illustrated in Figure 1; the architecture consists of six (6) stages as depicted in the figure. The architecture stages are described in the following sections.

2.1 Hotspot activation

The application has to be installed on the Internet provider and the consumer smartphones. The user can choose to connect to the application as a provider or a consumer, as shown in Figure 2.

In the provider side: (1) the Hotspot and Bluetooth are automatically activated and (2) different amounts of data are specified with their prices, i.e. 500 kb for a predefined charge of 5 Riyals.

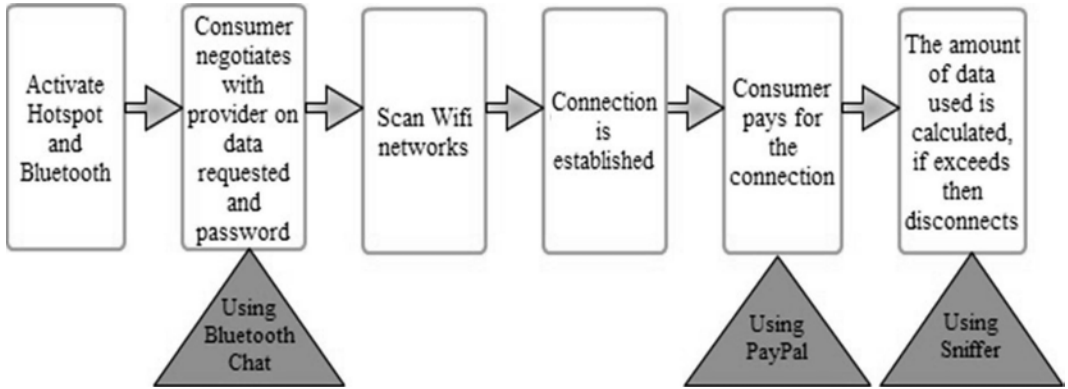


Figure 1. The application architecture.



Figure 2. The home interface.

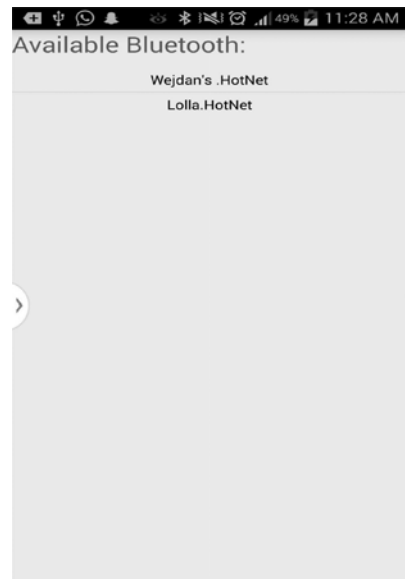


Figure 3. Scan of available Bluetooth devices.

And in the consumer side the Wi-Fi and Bluetooth are automatically activated.

2.2 Negotiation

The consumer scans the available Bluetooth devices in range, as shown in Figure 3, chooses a provider's Bluetooth and starts a negotiation with the provider by chatting via Bluetooth, as shown in Figure 4. If the negotiation is successful, the provider sends the password to the consumer.

2.3 Scan Wi-Fi networks

The application gives to the provider's Wi-Fi the same name as the provider's Bluetooth name, so the consumer can recognize the Wi-Fi name after

scanning the available Wi-Fi networks. Then the consumer connects to the desired Wi-Fi network with the password he received at the end of the negotiation stage. To distinguish the users of Hot-Net application, the text "HotNet" is added to mobile Hotspot name.

2.4 Connection and payment

After entering the password, the connection is established. Then the consumer pays for the amount of data he has chosen, the payment is performed using PayPal¹ Mock environment. PayPal

¹<https://www.paypal.com>.

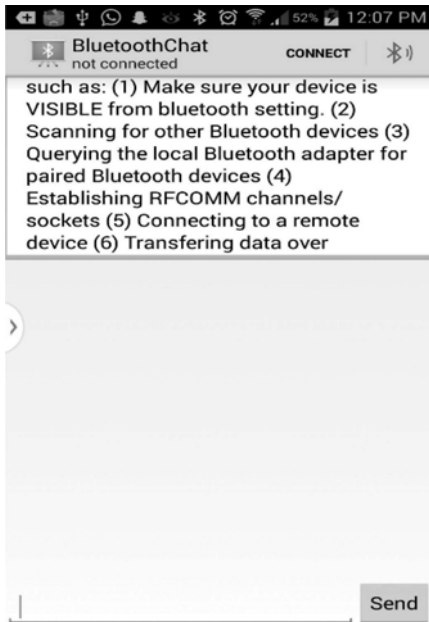


Figure 4. Bluetooth chat interface.

is a global leader in online payments. The goal is to offer a safe, secure, and reliable environment for users.

PayPal offers a set of Application Programming Interfaces (APIs) that give the means to incorporate PayPal functionality into a website applications and mobile apps. PayPal provides three environments that support calls to their API operations:

1. Mock environment: allows pre-packaged fake transactions only. The SDK will not attempt to contact PayPal's servers with this environment.
2. Sandbox environment: it is a virtual testing environment where fake user accounts are utilized to make calls to the PayPal operations without affecting any real PayPal users or their live PayPal accounts.
3. Live environment: in this environment, live accounts are owned by real people and containing real money.

2.5 Calculation of the used Internet data

The amount of Internet data consumed is calculated using a Sniffer, and the consumer is notified when he/she is close to reach the limit of the purchased data. The Wi-Fi connection is turned off when the amount of data is reached. A sniffer is a software program or a hardware device that monitors and analyzes the network traffic [10]. The sniffer examines the network traffic and

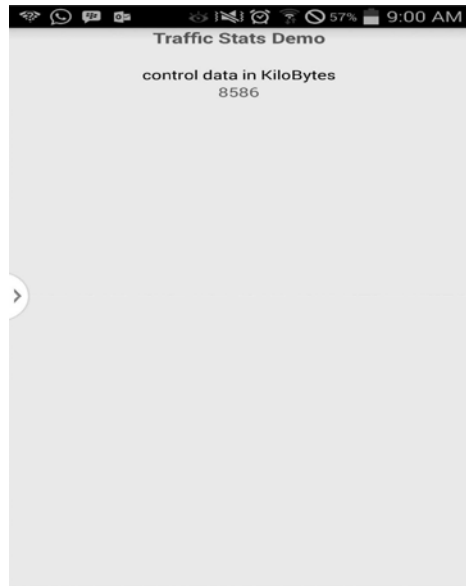


Figure 5. Sniffer interface.

makes a copy of the data. Actually the data is composed of packets; where a packet is the unit of data that is routed between an origin and a destination on the Internet. Sniffers are also sometimes called "network probes" or "snoops". The sniffer interface is shown in Figure 5.

3 EVALUATION

All the application functions were fully tested; test description and results are gives in the following.

3.1 Testing the Hotspot activation

There are two cases when testing the Hotspot activation (1) when the Wi-Fi is turned on; the application disabled the Wi-Fi and the mobile Hotspot was successfully turned on, (2) when Wi-Fi is turned off, the mobile Hotspot was successfully turned on.

3.2 Testing the search for available Hotspots

In order to test the Wi-Fi scan function, testers first scan all available networks, there were 5 networks and all of them appeared correctly. The second step is to limit the scanning result of the function to only hotspot networks using the application which ends with ".HotNet", 3 networks of 5 were Hotspot networks. The result of the test was successful, the program works correctly and only desirable networks appear.

3.3 Bluetooth testing

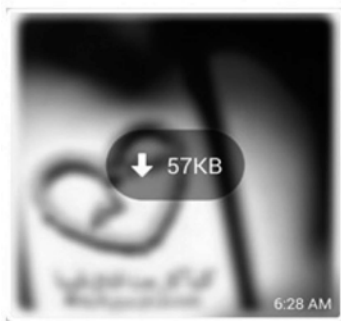
Two devices have been used in searching, pairing and sending string via Bluetooth, several password strings have been sent and they were received by the other user correctly.

3.4 Sniffer testing

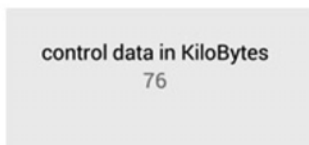
In order to ensure that the sniffer code is calculating correctly the used amount of data in the consumer side, we tested the code with different types of media with different sizes. We connected a mobile phone into a Wi-Fi Hotspot, then:

1. We downloaded an image with a size of 57 KB, as shown in Figure 6.a, and the size calculated by the sniffer is shown in Figure 6.b.
2. We downloaded a video with a size of 601 Kb, as shown in Figure 7.a. The amount of data calculated by the sniffer is 672 KB, as shown in Figure 7.b.
3. We downloaded an application from a play store, as shown in Figure 8.a. The amount of data calculated by the sniffer is shown in Figure 8.b.

The Figures 6–8 illustrate examples of sniffer code testing, more files with different types and sizes have been used in the testing.



(a)

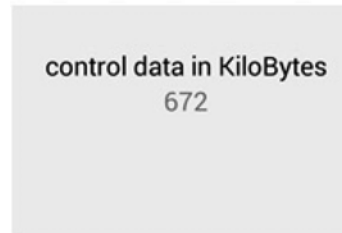


(b)

Figure 6. (a) The image size and (b) Size calculated by sniffer for the downloaded image.



(a)

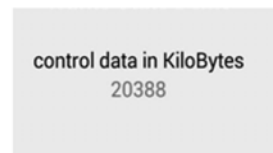


(b)

Figure 7. (a) The video size and (b) Size calculated by sniffer for downloaded video.



(a)



(b)

Figure 8. (a) The application size and (b) Size calculated by the sniffer for the downloaded application.

The size displayed by the sniffer is a little bit greater than the original size of the downloaded media file. The reason behind this is that the file is first segmented at the transport layer before transmission. Then, each segment will go through an encapsulation process resulting in a total amount of overhead equal to 66 bytes (20 bytes for TCP header + 20 bytes for IP + 26 bytes for Ethernet header). In other words, each segment; e.g. of size 536 bytes is incremented by at least 66 bytes resulting in total frame size of 602 bytes.

4 CONCLUSION AND FUTURE WORK

A mobile application architecture for selling the Internet data using the mobile Hotspot has been proposed in this paper. The user can use his smartphone as an Internet access point using the mobile Hotspot, share his mobile Internet connection and sell his Internet data. The proposed application can be considered as a prototype that can be improved. All the stages of the proposed application have been tested and all tests are successful. As future work, it will be interesting to improve the payment process in a way that the payment is not directly transferred to the provider, instead it is directed to a server and when the connection is turned off, the payment will be done in accordance of what amount of data the consumer actually used; in case the provider stops, for any reason, the connection earlier before the consumer receives the totality amount of requested data.

REFERENCES

- [1] Heer, T., Li, S., Wehrle, K. 2007. PISA: P2P Wi-Fi Internet Sharing Architecture. *In the 7th IEEE International Conference on Peer-to-Peer Computing, Los Alamitos.*
- [2] Sathiaselalan, A., Rotsos, C., Sriram C.S., Trossen, D., Papadimitriou, P., and Crowcroft, J. 2013. Virtual public networks. *In Proceedings of the 2013 Second European Workshop on Software Defined Networks EWSDN'13. Berlin, Germany, 10–11 October.*
- [3] Sastry, N., Crowcroft, J., Sollins, K., 2007. Architecting citywide ubiquitous Wi-Fi access. *In the Proceedings of the Sixth Workshop on Hot Topics in Networks. Atlanta, GA, USA, 14–15 November.*
- [4] Wong, M., Clement, A. 2007. Sharing Wireless Internet in Urban Neighbourhoods. *In the Proceedings of the Third Communities and Technologies Conference, Michigan State University.*
- [5] Heer, T., Gotz, S., Weingaertner, E., and Wehrle, K. 2008. Secure Wi-Fi Sharing at Global Scales. *In the Proceeding of 15th International Conference on Telecommunication (ICT). St. Petersburg, 16–19 June.*
- [6] Heer, T., Jansen, T., Hummen, R., Wirtz, H., Gotz, S., Weingaertner, E., and Wehrle, K. 2010. PiSA-SA: Municipal Wi-Fi Based on Wi-Fi Sharing. *In International Conference on Computer Communication Networks, ICCCN, Zurich, 2–5 August.*
- [7] Hummen, R., Wirtz, H., Viol, N., Heer, T., Wehrle, K. 2011. PISA-SA-security and mobility in a collaborative munifi. *Mobile Computing and Communications Review 15(3): 35–36.*
- [8] Wirtz, H., Hummen, R., Viol, N., Heer, T., Alejandra, M., Girón, L. and Wehrle, K. 2011. Cooperative Wi-Fi-sharing: Encouraging fair play. *In the Proceeding of ITU Kaleidoscope Academic Conference. Cape Town, South Africa, 12–14 December.*
- [9] Cao, Z., Fitschen, J. and Papadimitriou, P. 2015. Social Wi-Fi: Hotspot Sharing with Online Friends. *In the Proceeding of the IEEE 26th International Symposium on Personal, Indoor and Mobile Radio Communications—(PIMRC): Services Applications and Business. Hong Kong, Aug. 30–Sept. 2.*
- [10] Bradley, M. 2014. What Is a Sniffer in Computer Networking? Available at: http://compnetworking.about.com/od/networksecurityprivacy/g/bldef_sniffer.htm.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Big data mining: A classification perspective

Nojod M. Alotaibi & Manal A. Abdullah

Faculty of Computing and Information Technology, King Abdulaziz University (KAU), Saudi Arabia

ABSTRACT: An unprecedented amount of data is being generated and recorded every day. Big data is the term used to describe such data which is difficult to process, manage and analyze patterns using traditional databases or data mining algorithms. Mining big data is currently one of the most critical emerging research areas. Big data Mining refers to the process of extracting useful knowledge from large datasets or streams of data. Due to enormity, high dimensionality, heterogeneous, and distributed nature of data, traditional techniques of data mining may be unsuitable to work with big data. As a result, there is a critical need to develop effective and efficient big data mining techniques. This paper explores the current use of supervised classification algorithms for the big data. It also compares between the protocols based on their advantages and limitations.

Keywords: Big data, knowledge discovery, Data mining, Big Data mining, Supervised classification

1 INTRODUCTION

With the fast development of Internet communication and collaboration, Internet of Things and Cloud Computing, large amounts of data have become increasingly available at significant volumes (petabytes or more). Such data comes from a wider variety of sources and formats including social networking interactions, web pages, click streams, online transaction, emails, videos, audios, images, posts, search queries, health records, science data, sensors, smart phones and their applications, and so on [1]. According to the 2014 IDC 'Digital Universe Study' [2], 130 exabytes (EB) of world's data were created and stored in 2005. The amount grew to 4.4 zettabytes (ZB). It is doubling in size every two years and is projected to grow to 44 ZB in 2020 [2]. In 2012, IBM estimated that 2.5 quintillion bytes of data were created daily [3].

The rapid growth in the amount of data led to constitute the big data phenomenon. Since 2004, the interest of search on "big data" in Worldwide has increased exponentially, according to Google Trends (see Figure 1) [4].

There are three characteristics used to define big data (also called, the 3V's of big data): volume as data keeps growing, variety as the type of data is diverse, and velocity as it is continuously arriving very fast into the systems [1].

Due to these characteristics, the existing traditional techniques and technologies do not have the ability to handle storage and processing of this data. Therefore, new technologies have been developed to manage this big data phenomenon. IDC

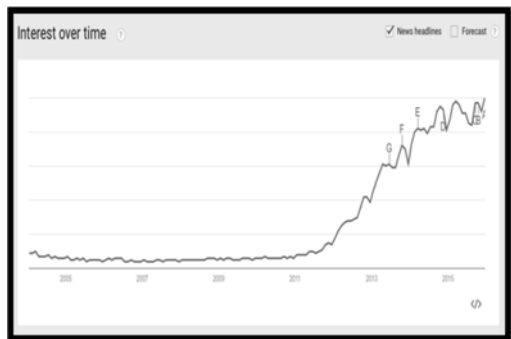


Figure 1. Worldwide interest: big data [4].

[5] defines big data technologies as "a new generation of technologies and architectures designed to extract value economically from very large volumes of a wide variety of data by enabling high velocity capture, discovery and analysis".

Mining and discovering meaningful knowledge from big data for decision-making, prediction, and for other purposes is extremely challenging due to its characteristics. Knowledge Discovery (KD) is the process of discovering useful knowledge from a collection of data. Major KD application areas include marketing, manufacturing, fraud detection, telecommunication, education, medical, Internet agent and many other areas [6, 7]. Data mining is the core step of KD process where algorithms are applied to extract useful patterns from data. Tasks in data mining can be classified into

clustering, classification, summarization, regression, association rule, sequence analysis, and dependency modeling.

Supervised classification is one of the most common tasks of data mining which concerned with prediction. The aim of the classification is to build a classifier based on the training data with known class labels to predict the class labels of new data [8]. There are various methods for data mining classification tasks such as: Decision tree's (TD), Support Vector Machine (SVM), genetic algorithms, neural networks, etc.

This paper is organized as follows. In Section 2, authors briefly review big data definitions and its related technologies. In Section 3, an overview of the KD and data mining is provided. Section 4 presents the concept of supervised classification. Big data mining and the related issues and challenges are described in Section 5. Section 6 explores some of current works of big data classification. Finally, authors give some conclusions in Section 7.

2 BIG DATA

In recent years, big data has become a hot research topic in many areas where storage and processing of massive amounts of data are required. In March 2012, American president Barack Obama administration announced the "Big Data Research and Development Initiative" with over \$200 million in research funding [9]. The goals of this initiative were to develop and improve technologies needed to collect, store, manage, and analyze this big data, to use these technologies to accelerate the pace of knowledge discovery in science and engineering fields, improve national security, and transform teaching and learning, and to expand the workforce required to develop and use big data technologies [9].

According to McKinsey [10] the term big data is used to refer to datasets whose size is beyond the capability of existing database software tools to capture, store, manage and analyze within a tolerable amount of time. However, there is no single definition of big data. O'Reilly [11] defines big data as "data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn't fit the structures of existing database architectures. To gain value from this data, there must be an alternative way to process it".

As seen from the above definitions, the volume of data is not the only characteristic of big data. In fact, big data has three major characteristics (known as 3V's), shown in Figure 2, which were first defined by Doug Laney in 2001 [12].

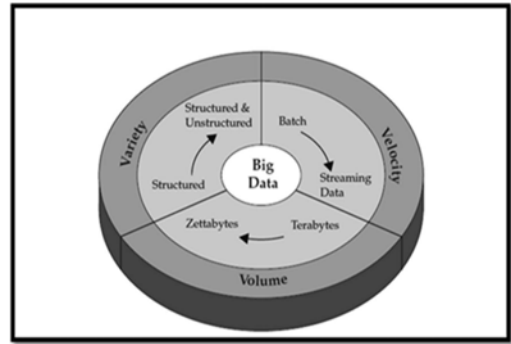


Figure 2. Three V's of big data [1].

- *Data volume* (i.e. the size of data) is the primary attribute of big data. The size of data could reach terabytes (TB, 10^{12} B), petabytes (PB, 10^{15} B), exabytes (EB, 10^{18} B), zettabytes (ZB, 10^{21} B) and more. For example, Facebook reached more than 8 billion video views per day in September 2015 [13].
- *Variety* refers to the fact that big data can come from different data sources in various formats and structures. These data sources are divided into three types: structured, semi-structured and unstructured data [14]. Structured data is described as data that follows a fixed schema. An example of this type is a relational database system. Semi-structured data is a type of structured data, but it doesn't have a rigid structure [15]. Its structure may change rapidly or unpredictably [15]. Examples include weblogs and social media feeds. Unstructured data refers to data that cannot be stored into relational tables for analysis and querying. This data represents 80% of the world's data. Files or documents such as videos, images, audio, PDF and spreadsheet are examples.
- The *velocity* of data refers to the increasing rate at which data flows into an organization [11].

More recently, two additional V's have been added to define big data: veracity and value. *Veracity* (uncertainty of data) refers to the accuracy, integrity, and quality of the data being collected, while *value* refers to the worth of the data being extracted [17].

All previous characteristics of big data are considered challenging and this the reason why we cannot use traditional Database Management Systems (DBMS) in the processing and analyzing big data. As a result, new technologies have been developed to meet the challenges. Following subsection discusses some of these technologies.

2.1 Big data technologies

Big Data is a new term used to identify the datasets that due to their large size and complexity, which cannot be managed with traditional database systems. In recent years, there are many technologies have been developed to process this huge volumes of data.

Apache Hadoop [18] is an open source software framework that enables the distributed processing of large data sets across clusters of commodity hardware using simple programming models. There are two main components of Hadoop: Hadoop Distributed File System (HDFS) and MapReduce. HDFS is a distributed, scalable file system written in java for the Hadoop framework. MapReduce is a programming paradigm that allows users to define two functions, map and reduce, to process large number data in parallel. Companies like Facebook, Yahoo!, Amazon, Baidu, AOL, and IBM use Hadoop on a daily basis. Hadoop has many advantages include [19]: cost effective, fault tolerant, flexibility, and scalability. Hadoop has many other related software projects that uses the MapReduce and HDFS framework such as Apache Pig, Apache Hive, Apache Mahout, Apache HBase, and others [18].

Apache Pig [1] was originally developed at Yahoo in 2006 for processing big data. In 2007, it was moved into the Apache Software Foundation. It allows people using Hadoop to focus more on analyzing large data sets and spend less time having to write MapReduce programs.

Apache Hive [20] was developed at Facebook in 2009. It is data warehouse software for querying and managing large datasets residing in distributed storage. It built on top of Apache Hadoop. Hive defines a simple SQL-like query language, called Hive Query Language (HQL), which enables users familiar with SQL to query the data. Hive is optimized for scalability, extensibility, and fault-tolerance.

Apache HBase [21] is a distributed columnar database that supports structured data storage for very large tables.

Jaql [22] was created by workers at IBM Research Labs in 2008 and released to open source. It is a query language for JavaScript Object Notation (JSON), but it supports more than just JSON such as XML, CSV, flat files, and more.

Storm [23] was created at Backtype, a company acquired by Twitter in 2011. It is a free and open source distributed real-time computation system that does for real-time processing what Hadoop does for batch processing. Storm offers features such as scalability, fault-tolerant, and distributed computation.

NoSQL Database [24] (Not only SQL) is a term used to designate database management systems that differ from classic RDBMS in some way.

These data stores may not require fixed table schemas, usually avoid join operations, do not attempt to provide ACID (atomicity, consistency, isolation, durability) properties and typically scale horizontally. There are several types of NoSQL database:

- *Key-value stores*. In key-value store, each single item in the database is stored as an attribute name (or key), together with its value. Examples of key-value store are Amazon's Dynamo and Oracle's BerkeleyDB.
- *Document-oriented database*. It is a database designed for storing, retrieving and managing document-oriented or semi-structured data. Examples of these databases are CouchDB and MongoDB.
- *Column stores*. It stores columns of data together, instead of rows. Examples include Cassandra and Apache HBase.
- *Graph database*. It contains nodes, edges and properties to represent and store data. Examples of graph databases are Neo4j and HyperGraphDB.

3 DATA MINING

Knowledge Discovery (KD) is the process of extracting useful knowledge from huge volumes of data. It can be defined as the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data [7]. The KD process consists of the following steps as shown in Figure 3 [7]:

1. *Understanding the application domain*. It includes learning prior knowledge and the user's goals
2. *Creating a target data set*. In this step, a subset of variables and data are selected that will be used to perform discovery task.
3. *Data cleaning and preprocessing*. It includes the basic operations: removing noise, dealing with missing values.
4. *Data reduction and projection*. It involves of finding useful attributes to represent data.
5. *Choosing the data mining task*. There are several data mining tasks include: clustering, classification, regression, summarization, etc.
6. *Choosing the data mining algorithms*. In this step, appropriate methods are selected to be used for searching for patterns in the data.
7. *Data mining*. Searching for patterns in a particular representational form (such as classification rules or trees, regression and clustering) using the selected data mining methods.
8. *Interpretation*. Interpreting mined patterns and possibly returns to any of the previous steps for further iteration if the pattern evaluated is not useful.

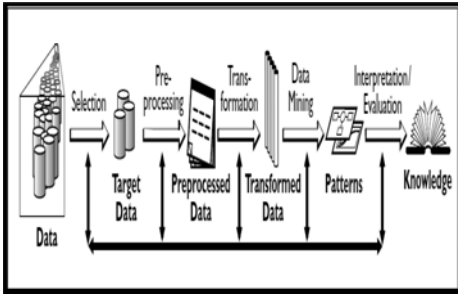


Figure 3. A typical knowledge discovery process [7].

9. *Using discovered knowledge.* The final step consists of incorporating the discovered knowledge into another system, or documenting and reporting it to interested parties.

Data mining is the core step in the whole KD process. It consists of applying data analysis and discovery algorithms and produce enumeration of patterns (models) over data. It is widely used in fields such as science, engineering, economics, social media, medicine, marketing, and business. Currently, many data mining tools are available for free on the Web such as Waikato Environment for Knowledge Analysis (WEKA) [25], RapidMiner [26], Orange [27], Konstanz Information Miner (KNIME) [28], and more. The major tasks of data mining can be classified as the following [3]:

- *Clustering:* maps a data item into one of several clusters, where clusters are natural grouping of data items based on similarity or probability density models.
- *Classification:* classifies a data item into one of several predefined categorical classes.
- *Regression:* maps a data item to real-valued prediction variable. It is used in different prediction and modeling applications.
- *Summarization:* provides compact description for a subset of data. Examples include mean and standard deviation of fields.
- *Association rule:* describes association relationship among different attributes.
- *Sequence analysis:* models sequential patterns, like time-series data. The goal is to model the process of generating the sequence or to extract and report deviation and trends over time.
- *Dependency modeling:* describes significant dependencies between variables.

4 CLASSIFICATION

Classification is one of the most common types of data mining, which finds patterns in information and

categorizes them into different classes. It is supervised learning, which generates a classifier (model) based on a set of instances with known labels, is called the training set. Then, the classifier is used for classifying new or previously unseen data [8]. The converse of this is unsupervised learning, which involves classifying data into categories based on some similarity of input parameters in the data. Examples of supervised classification are spam detection, credit card fraud detection, and medical diagnosis. There are three different types of supervised classification: binary, multi-class, and multi-label classification. In binary classification, each instance of data may belong to one of two possible class labels. In multi-class classification, more than two class labels are involved and each instance is assigned to only one class label. In case of multi-label classification, there are more than two class labels and each instance may belong to more than one class label at same time. There are many types of classification algorithms for extracting knowledge from data, which can be categorized into: logic-based techniques (C4.5, CART, and RIPPER), perceptron-based techniques (Artificial Neural Networks), statistical learning techniques (Naive Bayes classifiers and Bayesian networks), and instance-based techniques (k-nearest neighbor) [8].

5 BIG DATA MINING

In the present age, huge amount of data are produced every moment in various fields such as science, Internet, and physical systems. This is big data. Useful knowledge can be extracted from this big data with the help of data mining. Due to enormity, high dimensionality, heterogeneous, and distributed nature of data, traditional techniques of data mining may be unsuitable for extracting knowledge from this data. Mining big data is an emerging research area, hence a plethora of possible future research directions arise. The objectives of big data mining techniques go beyond fetching the requested information or even uncovering some hidden relationships and patterns [29]. Comparing with the results derived from mining the traditional datasets, unveiling the massive volume of interconnected heterogeneous big data has the potential to maximize our knowledge and insights in the target domain. Begoli and Horey in [30] proposed three principles for effective knowledge discovery from big data: first, the architecture should support many analysis methods such as data mining, statistical analysis, machine learning, and visualization. Second, different storage mechanism should be used because all data cannot fit in a single storage. Also, the data should be stored and processed at all stages of the pipeline. Third, the results should be accessible and easy to understand.

5.1 Issues and challenges of big data mining

There are a number of issues and challenges related to big data mining as follows [31]:

- *Heterogeneity or variety*: an existing data mining techniques have been used to discover unknown patterns and relationships of interest from structured, homogeneous, and small datasets. Variety is one of the fundamental characteristics of big data, comes from the phenomenon that there exists unlimited different sources that generates and contributes to big data. The data from different data sources may formed interconnected, interrelated, and delicately and inconsistently represented data. Mining useful information from such data is great challenge. Heterogeneity in big data also means that it is an obligation to accept and deal with structured, semi-structured, and even entirely unstructured data concurrently.
- *Scalability or volume*: the extraordinary volume requires high scalability of its data management and mining tools. Cloud computing with parallelism can deal with the volume challenge of big data.
- *Speed or velocity*: the ability of fast accessing and mining big data is highly essential-processing/mining of task must be completed within a definite period of time, otherwise, the results becomes less valuable or even worthless.
- *Accuracy and trust*: with big data, the data sources are of many different origins, not all well-known, and not all confirmable. As a result, the accuracy and trust of the source data quickly become a serious concern.
- *Privacy crisis*: data privacy has been always a challenge. The concern has become extremely serious with big data mining that often requires personal information in order to produce relevant/accurate results such as location-based and personalized services. Additionally, with the enormous volume of big data such as social media that contains tremendous amount of highly interconnected personal information. When all bits of information about a person are dug out and put together, any privacy about that individual instantly disappears.
- *Interactiveness*: it means the capability of a data mining system that allows fast and adequate user interaction such as feedback/interference/guidance from users. It relates to all the characteristics of big data and can help overcome the challenges coming along with each of them.

6 BIG DATA CLASSIFICATION

Because of big data characteristics, traditional data mining algorithms may not be suitable to

mine such huge data. As a consequence, there is an urgent need for developing algorithms and techniques capable of mining big data while dealing with their inherent properties. Several studies attempted to improve the traditional classification algorithms to make them work with big data, to parallelize classification algorithms based on MapReduce or to develop new software tools to mining big data. The approaches of big data classification are summarized in Figure 4.

6.1 Improving traditional classification algorithms

Niu et al. [32] improved the traditional KNN algorithm and proposed a new algorithm, called Neighbor Filter Classification (NFC) to realize fast classification operation in big data. Lui [33] proposed a new improved model for the original random forest algorithm in big data environment. The proposed model has higher classification accuracy.

Support Vector Machines (SVMs) is one of the most popular techniques for data classification and regression. Their computation and storage requirements increase rapidly with the size of the dataset, making it unsuitable for big data. Many researchers have tried to find possible methods to apply SVM classification for large data sets. Rebentrost et al. [34] presented a quantum-based support vector machine algorithm for big data classification. The proposed algorithm can achieve exponential speedup over classical algorithm. In [35] Cervantes et al. presented SVM for classification large data using minimum enclosing ball clustering. The proposed approach has good classification accuracy compared with classical SVM. Cervantes et al. [36]

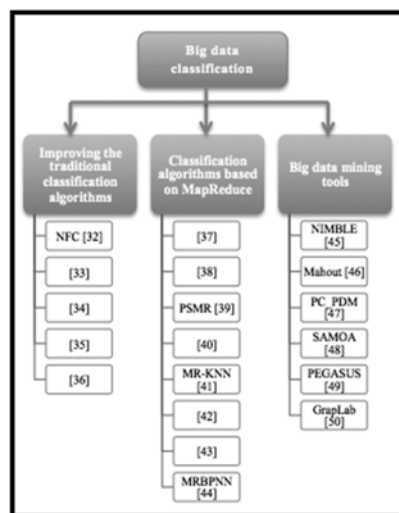


Figure 4. Big data classification approaches.

Table 1. Summarization of big data classification algorithms.

Reference number	Algorithm	Limitations	Modification	Advantages
[32]	KNN	<ul style="list-style-type: none"> – Time cost of modeling is unacceptable. – Sensitive to parameter K. 	Neighbor Filter Classification (NFC).	<ul style="list-style-type: none"> – It reduces the computational cost to $O(n)$. – It is able to replace or adjust key input parameters automatically. – It updates other parameters regularly.
[33]	Random forest	<ul style="list-style-type: none"> – Accuracy of a random forest will gradually reduce over time. 	Improved random forest.	<ul style="list-style-type: none"> – It has higher classification accuracy than the traditional random forest.
[34]	SVM	<ul style="list-style-type: none"> – High computational complexity (long training time) and extensive memory requirements of the required quadratic programming in large-scale tasks. 	Quantum least-squares SVM.	<ul style="list-style-type: none"> – Achieve exponential speedup over classical algorithm: $O(\log NM)$ in both training and classification stages.
[35]			SVM using Minimum Enclosing Ball (MEB) clustering.	<ul style="list-style-type: none"> – It provides good classification accuracy compared with classic SVM, while the training time is significantly shorter.
[36]			SVM based on fuzzy clustering.	<ul style="list-style-type: none"> – It achieves good performance for large datasets and fast convergence speed.
[37]	NBC	<ul style="list-style-type: none"> – It does not scale up well when the dataset is large. 	Implementing NBC on top of Hadoop MapReduce framework.	<ul style="list-style-type: none"> – The accuracy of NBC is improved and approaches 82% when the dataset size increase.
[38]	SVM	<ul style="list-style-type: none"> – The computation and storage requirement increases tremendously for large dataset. 	MapReduce based parallel SVM algorithm.	<ul style="list-style-type: none"> – It works efficiently on large datasets as compared to the sequential SVM. – The computation time taken by the SVM with multi-node cluster is less as compared to the single node cluster for large dataset.
[39]			Parallel SVM based on MapReduce (PSMR).	<ul style="list-style-type: none"> – The training time is reduced significantly.
[40]			Ontology enhanced parallel SVM based on MapReduce.	<ul style="list-style-type: none"> – It reduces the training time significantly.
[41]	KNN	<ul style="list-style-type: none"> – The complexity of KNN is $O(n \cdot D)$, where n is the number of instances and D the number of features. – Memory consumption problems. 	MapReduce-based K-Nearest Neighbor approach (MR-KNN).	<ul style="list-style-type: none"> – The reduction of computational time achieved compared to the utilization of the sequential version.
[42]	KNN-join	<ul style="list-style-type: none"> – It needs to spend a lot of time to handle large volume data 	Parallel MapReduce based KNN-join.	<ul style="list-style-type: none"> – It achieves higher performance than the serial one.
[43]	C4.5	<ul style="list-style-type: none"> – The process of building decision trees can be very time consuming when the dataset is extremely big. 	Parallel C4.5 decision tree classification algorithm based on MapReduce.	<ul style="list-style-type: none"> – It exhibits both time efficiency and scalability.
[44]	Back-propagation neural network (PBNN)	<ul style="list-style-type: none"> – The computation process of ANN is slow especially when dealing with large datasets. 	MapReduce based parallel back-propagation neural network (MRBPNN).	<ul style="list-style-type: none"> – The computation overhead of neural network can be significantly reduced.

proposed an SVM classification algorithm based on fuzzy clustering. The proposed approach is scalable to large data sets with high classification accuracy and fast convergence speed.

6.2 Classification algorithms based on MapReduce

Liu et al. [37] designed big data analyzing system to classify millions of movie reviews using a Naïve Bayes Classifier (NBC). They implemented the NBC on top of Hadoop framework, with some additional modules. The results show that the accuracy of NBC is improved and approaches 82% when the dataset size increases. Priyadarshini and Agarwal [38] proposed a MapReduce based parallel SVM algorithm for big data classification. The software used was lib-SVM. In the proposed algorithm, the training data is divided into subsets and each subset is trained with SVM. Then, the support vectors of two SVMs are combined to be input for the next SVM. This process is repeated until only one set of support vectors is left. Xu et al. [39] proposed a parallel SVM based on MapReduce (PSMR) algorithm for email classification. The parallel SVM is based on the cascade SVM model. Caruana et al. [40] developed a new algorithm for parallelized SVM based on MapReduce framework for scalable spam filter training. The parallel SVM is built on the Sequential Minimal Optimization (SMO) algorithm. Ontology semantics are used to minimize the accuracy degradation when distributing the training data among a number of SVM classifiers. Maillou et al. [41] proposed a MapReduce-based K-Nearest Neighbor (KNN) approach (MR-KNN) for big data classification. Yan et al. [42] proposed a parallel KNN-join algorithm using MapReduce for big data multi-label classification. Dai and Ji [43] suggested a parallel C4.5 decision tree classification algorithm based on MapReduce. Liu et al. [44] proposed a MapReduce Based Parallel Back-Propagation Neural Network (MRBPNN). In this work, three parallel neural networks are presented to deal with data intensive scenarios in terms of the volume of classification data, the size of the training data, and the number of neurons in NN. They concluded that the computation overhead of NN can be significantly reduced using number of computers in parallel.

These big data classification algorithms with their advantages and limitations are summarized in Table 1.

6.3 Big data mining tools

In big data mining, there are many open source tools. Some of these tools are summarized in the following:

NIMBLE [45] is a portable infrastructure that enables rapid development of parallel machine learning and data mining algorithms. It runs on the top of Hadoop framework.

Apache Mahout [46] is open source project by Apache Software Foundation (ASF). Mahout is written in java and provides scalable data mining algorithms. It contains implementations for clustering, categorization, Collaborative Filtering (CF), and evolutionary programming on top of Apache Hadoop.

Big Cloud-Parallel Data Mining (BC-PDM) [47] is a cloud-based data mining platform that provides access to large telecom data and business solutions for telecom operators. BC-PDM is based on the MapReduce implementation of cloud computing. It supports parallel ETL process (extract, transform, and load), statistical analysis, data mining, text mining, and social network analysis.

Apache SAMOA (Scalable Advanced Massive Online Analysis) [48] is a platform for mining big data streams. It includes distributed algorithms for common machine learning tasks.

PEGASUS (Peta-Scale Graph Mining System) [49] which is a graph mining system for very large graphs built on top of the Hadoop framework.

GraphLab [50] is high-level graph-parallel system built without using MapReduce. It is an open source project written in C++.

7 CONCLUSION

Big data has become a hot research topic that attracts extensive attention from academia, industry, and governments around the world. In this paper, we briefly introduce the concept of big data, including its definitions, characteristics, and technologies. This paper also provides an overview of big data mining and discuss the related issues and challenges. To support big data mining, we briefly describe the overview of supervised classification algorithms over big data.

REFERENCES

- [1] Zikopoulos P., Eaton C., deRoos D., Deutsch T., and Lapis G., *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*, 1 s ed., Sit S., Ed. USA: McGraw-Hill Companies, 2012.
- [2] Tuner V., Reinsel D., Gantz J., and Minton S., "The Digital Universe of Opportunities: Rich Data and The Increasing Value of The Internet of Things," EMC Corporation, Apr. 2014.
- [3] (2013) What is Big Data?: Bringing Big Data to The Enterprise. [Online]. Available: <https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>.
- [4] (2015) Google Trends. [Online]. Available: <http://www.google.com/trends/explore#q=big%20data>.

- [5] Gantz J. and Reinsel D., "The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East," EMC Corporation, Dec. 2012.
- [6] Singh P., Gosawi G., and Dubey S., "Application of Data Mining," *Binary Journal of Data Mining and Networking*, vol. 4, pp. 41–44, 2014.
- [7] Fayyad U., Piatetsky-Shapiro G., and Smyth P., "The KDD Process for Extracting Useful Knowledge from Volumes of Data," *Communications of the ACM*, vol. 39, no. 11, pp. 27–34, Nov. 1996.
- [8] Kotsiantis S., "Supervised Machine Learning: A Review of Classification Techniques," *Informatica*, vol. 31, pp. 249–268, July 2007.
- [9] (2012) Obama Administration Unveil "Big Data" Initiative: Announces \$200 Million in New R&D Investments. [Online]. Available: https://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release.pdf.
- [10] Manyika J., Chui M., Brown B., Bughin J., Dobbs R., Roxburgh C., and Byers A., "Big Data: The next frontier for innovation, competition and productivity," McKinsey Global Institute, May 2011.
- [11] Dumbill E., Croll A., Steele J., and Loukides M., *Planning for big data*, Beijing: O'Reilly Media, 2012.
- [12] Laney D., "3D Data Management: Controlling Data Volume, Velocity and Variety," META Group Inc., Feb. 2001.
- [13] (2015) Facebook Reports Third Quarter 2015 Results. [Online]. Available: <http://www.techmeme.com/151104/p24#a151104p24>.
- [14] Sagiroglu S. and Sinanc D., "Big Data: A Review," in *Proc. of the 2013 International Conference on Collaboration Technologies and Systems (CTS)*, 2013, pp. 42–47.
- [15] Pankowski T., "Querying Semistructured Data Using a Rule-Oriented XML Query Language," in *Proc. of the 15th European Conference on Artificial Intelligence (ECAI)*, 2002, pp. 302–206.
- [16] Beyer M. and Laney D., "The Importance of 'Big Data': A Definition," Gartner, 2012.
- [17] Hassanién A., Azar A., Snasel V., Kacprzyk J., and Abawajy J., *Big Data in Complex Systems: Challenges and Opportunities*, 1s ed., Kacprzyk J., Ed. Springer International Publishing, 2015.
- [18] (2014) Apache Hadoop. [Online]. Available: <https://hadoop.apache.org/>.
- [19] Mirajkar N., Bhujal S., and Deshmukh A., "Perform wordcount Map-Reduce Job in Single Node Apache Hadoop Cluster and Compress Data Using Lempel-Ziv-Oberhumer (LZO) algorithm," *International Journal of Computing Science Issues (IJCSI)*, vol. 10, pp. 719–728, Jan. 2013.
- [20] (2014) Apache Hive. [Online]. Available: <http://hive.apache.org/>.
- [21] (2015) Apache HBase. [Online]. Available: <http://hbase.apache.org/>.
- [22] What is Jaql? [Online]. Available: <http://www-01.ibm.com/software/data/infosphere/hadoop/jaql/>.
- [23] (2015) Apache Storm. [Online]. Available: <http://storm-project.net/>.
- [24] (2015) What is NoSQL? [Online]. Available: <https://www.mongodb.com/nosql-explained>.
- [25] (2015) WEKA: The University of Waikato. [Online]. Available: <http://www.cs.waikato.ac.nz/ml/weka/>.
- [26] (2015) RapidMiner. [Online]. Available: <https://rapidminer.com/>.
- [27] (2015) Orange: Data Mining Fruitful and Fun. [Online]. Available: <http://orange.biolab.si/>.
- [28] (2015) KNIME. [Online]. Available: <https://www.knime.org/>.
- [29] Prakash B. and Hanumanthappa M., "Issues and Challenges in the Era of Big Data Mining," *International Journal of Emerging Trends and Technology in Computer Science (IJETICS)*, vol. 3, pp. 321–325, 2014.
- [30] Begoli E. and Horey J., "Design Principles for Effective Knowledge Discovery From Big Data," in *Proc. of the Joint Working IEEE/IFIP Conference on Software Architecture (WICSA) and European Conference on Software Architecture (ECSA)*, 2012, pp. 215–218.
- [31] Hong B., Meng X., Chen L., Winiwarer W., and Song W., *Database Systems for Advanced Applications*, 1s ed., Berlin: Springer, 2013.
- [32] Niu K., Zhao F., and Zhang S., "A Fast Classification Algorithm for Big Data Based on KNN," *Journal of Applied Science*, vol. 13, pp. 2208–2212, 2013.
- [33] Lui Y., "Random Forest Algorithm in Big Data Environment," *Computer Modelling & New Technologies*, vol. 18, pp. 147–151, 2014.
- [34] Rebentrost P., Mohseni M., and Lloyd S., "Quantum Support Vector Machine for Big Data Classification," *Physical review letters*, vol. 113, pp. 1–5, Sept. 2014.
- [35] Cervantes J., Li X., Yu W., and Li K., "Support vector machine classification for large data sets via minimum enclosing ball clustering," *Neurocomputing*, vol. 71, pp. 611–619, 2008.
- [36] Cervantes J., Li X., and Yu W., "Support Vector Machine Classification Based on Fuzzy Clustering for Large Data Sets," in *Proc. of The 5th Mexican International Conference on Artificial Intelligence (MICAI)*, 2006, pp. 572–582.
- [37] Liu B., Blasch E., Chen Y., Shen D., and Chen G., "Scalable Sentiment Classification for Big Data Analysis Using Naïve Bayes Classifier," in *Proc. of The 2013 IEEE International Conference on Big Data*, 2013, pp. 99–104.
- [38] Priyadarshini A. and Agarwal S., "A Map Reduce based Support Vector Machine for Big Data Classification," *International Journal of Database Theory and Application*, vol. 8, pp. 77–98, 2015.
- [39] Xu K., Wen C., Yuan Q., He X., and Tie J., "A MapReduce based Parallel SVM for Email Classification," *Journal of Networks*, vol. 9, pp. 1640–1647, June 2014.
- [40] Caruana G., Li M., and Liu Y., "An Ontology Enhanced Parallel SVM for Scalable Spam Filter Training," *Journal of Neurocomputing*, vol. 108, pp. 45–57, May 2013.
- [41] Mailló J., Triguero I., and Herrera F., "A MapReduce-based k-Nearest Neighbor Approach for Big Data Classification," in *Proc. of The 2015 IEEE Trustcom/BigDataSE/ISPA Conference*, 2015, pp. 167–172.

- [42] Yan X., Wang Z., Zeng D., Hu C., and Yao H., "Design and Analysis of Parallel MapReduce based KNN-join Algorithm for Big Data Classification," *TELKOMNIKA Indonesian Journal of Electrical Engineering*, vol. 12, pp. 7927–7934, Nov. 2014.
- [43] Dai W. and Ji W., "A MapReduce Implementation of C4.5 Decision Tree Algorithm," *International Journal of Database Theory and Application*, vol. 7, pp. 49–60, 2014.
- [44] Liu Y., Yang J., Huang Y., Xu L., Li S., and Qi M., "MapReduce Based Parallel Neural Networks in Enabling Large Scale Machine Learning," *Computational Intelligence and Neuroscience*, vol. 2015, pp. 1–13, Aug. 2015.
- [45] Ghoting A., Kambadur P., Pednault E., and Kannan R., "NIMBLE: a Toolkit for the Implementation of Parallel Data Mining and Machine Learning Algorithms on MapReduce," in *Proc. of The 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2011, pp. 334–342.
- [46] (2009) Introducing Apache Mahout. [Online]. Available: <http://www.ibm.com/developerworks/library/j-mahout/>.
- [47] Yu L., Zheng J., Wu B., and Wang B., "BC-PDM: Data Mining, Social Network Analysis and Text Mining System Based on Cloud Computing," in *Proc. of The 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2012, pp. 1496–1499.
- [48] Morales G. and Bifet A., "SAMOA: Scalable Advanced Massive Online Analysis," *Journal of Machine Learning Research*, vol. 16, pp. 149–153, 2015.
- [49] Kang U., Tsourakakis C., and Faloutsos C., "PEGASUS: A Peta-Scale Graph Mining System Implementation and Observations," in *Proc. of The 9th IEEE International Conference on Data Mining*, 2009, pp. 229–238.
- [50] Low Y., Bickson D., Gonzalez J., Guestrin C., Kyrola A., and Hellerstein, J., "Distributed GraphLab: A Framework for Machine Learning and Data Mining in The Cloud," *Journal of VLDB Endowment*, vol. 5, pp. 716–727, Apr. 2012.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

ICT drivers of intelligent enterprises

Monika Łobaziewicz

University College of Enterprise and Administration in Lublin, Lublin, Poland

ABSTRACT: In organizations in every branch, managers wonder whether they are getting full value from the massive amounts of data and information they already have within their organizations. Systems and tools based on the ICT are collecting more data than ever before, yet many enterprises are still looking for better ways to obtain value from their data and compete in the marketplace. Managers now need businesses run on data-driven decisions and use optimal solutions based on complex business parameters, and they have to take right actions quickly because of market and business partners pressure. IT technology and instrumentation for intelligent connections are abundantly available, and at a relatively low cost, but to turn this information into new intelligent action, organizations need knowledge and people who are able to put information into context, and use them in a business process management. Therefore, there are many discussions related to an intelligent organization. This paper deals with ICT drivers of intelligent enterprises based on the discussion related to the research conducted by the government institution in Poland, by MIT Sloan Management Review partnered with the IBM Institute for Business Value conducted in different countries and the own survey.

1 INTRODUCTION

With the evolution of information technologies, systems, applications, professional IT tools to drill data generated in organizations, a discussion about intelligent enterprise has been going on. The way which and how data and information are processed is vital for any significant improvements in an organization today. Without the knowledge, a skill of using data or information to create an organization value, it is difficult to compete in the marketplace. Since a few years, the term “intelligent” has got a special meaning, next to the term “innovation”. Thus, the enterprise intelligence may be considered as a methodological structure institutionalizing the adaptation through the continuous adjustment and evaluation. This allows the company to deal with unknown market situations, to adapt to new approaches, strategic and tactical concepts, and to develop key competences. It means an “engine of innovation” allowing for a constantly company repositioning providing in this way the desired competitiveness on the dynamically changing market (Thannhuber, 2005).

The aim of this article is the identification of ICT drivers empowering the intelligent enterprise based on the confrontation of the research conducted in Poland and other countries. The paper consists of three parts. The first chapter is dedicated for a concept of intelligent enterprise. In the second section, there were presented final results of research conducted by Polish Agency for Enter-

prise Development and MIT Sloan Management Review partnered with the IBM Institute for Business Value noticed ICT drivers empowering intelligent organizations. The third part presents ICT drivers pointed as a result of own surveys made in a group of enterprises that the author collaborate with. This is the first approach to the problem.

2 THE CONCEPT OF AN INTELLIGENT ENTERPRISE

The concept of an intelligent enterprise has its source in the science of management and particularly is correlated with the following ideas: an organisation as an information system, a learning organization, knowledge management or an intellectual capital of an organisation. A wide scope of theoretical basics causes an immensity of approaches and definitions, and then, in further consequence, the lack of a universal definition that would be commonly acceptable.

An intelligent organisation is able to collect, process, interpret and communicate information needed for decision-making processes (Wilensky, 1967).

I. Nonaka and H. Takeuchi believe that an intelligent organization is not based on specialized R & D departments, but is based on its members, the way in which they behave and the culture they present, as part of which each is an employee and entrepreneur with his or her knowledge brought in (Nonaka, Takeuchi, 1995).

Intelligent organisations constantly learn. This process of learning consists of an observation of an external and internal environment, development of the perception of understanding the environment, giving the meaning through an interpretation and undertaking activities and improvement of organisation behaviour (Hamel, Prahalad, 1994; Muryjas, Wawer, 2014).

An intelligent organisation has a capacity to teach not to do things which are just old habits or routine. It is able to change those activities that bring no progress or are wrong (Christensen, 1997).

Therefore, an intelligent enterprise has some special abilities that distinguish it from other:

- adaptability to a changing environment;
- ability to influence and shape the environment;
- ability to find new strategic domains (in the product-market system) in their external environment and the rapid reconfiguration of resources according to the new domain;
- ability to make a positively co-contribute to the development of their environment in the context of sustainable development (Schwaninger, 2009).

The business model of a modern enterprise with the characteristics of an intelligent organization is fundamentally different from the traditional business model. The most important features that distinguish the two types of models include:

- the transition from focusing on improving production processes to focusing on innovation as an essential factor of value creation; human capital, a creative factor of the development of intelligent organization is its key value;
- the transition from high to low states of working human capital by optimizing supply chain management systems and customer relationship management (from the point of view of information systems corresponding class applications include Enterprise Resource Planning—ERP and Customer Resource Management—CRM);
- the transition from high to low states of physical capital, and recognition of intellectual capital as a fundamental value creation medium (Davenport, Leibold, Voelpel, 2006).

The above described features of the business model of an intelligent organization cause a shift of attention from the improvement of existing processes as a development factor to an ongoing changes of existing processes through the mechanisms of a learning organization. In the intelligent organization an attention is moved in the management of enterprises from material resources to intangible resources that require specific management competences for a continuous process of converting information into intelligence.

3 ICT DRIVERS OF INTELLIGENT ORGANIZATIONS IN THE RESEARCH

First research, that was carried out in Poland, concerning intelligent organisations was made by Polish Agency for Enterprise Development (PARP) in 2010 on a group of 300 enterprises. The purpose of the research was finding an answer to a question whether small and medium-sized enterprises in Poland use the solutions dedicated for intelligent organisations and whether activities they undertake improve competitiveness. Poland is at the stage of an intensive development and investing in the research—development sector, as well as the introduction of product and process innovations based on ICT, dominated by intelligent solutions, therefore, no new research in this field have been recently carried out. In the research carried out by PARP it was assumed that an intelligent organisation fulfils at least four conditions specified below:

- it has a strategy of development formed which includes long-term goals to achieve and possible ways of their achievement;
- it has a personnel management policy well-developed;
- it has a company website and intra network as well as it uses specialised computer software
- during the process of making a purchase or sale it exchanges knowledge with the environment in other way

Studies have shown that in the SME sector 26.5% of the participating companies had a strategy of development formed, 31.6% of them had the personnel management policy, 47% a developed computer software and 38% exchanged knowledge with the environment during the process of making a purchase or sale. However, the results indicated that almost 63% of companies which belonged to the sector of large enterprises have both the strategy of development policy and the personnel management policy well formed.

Therefore, the results incline to observe that bigger organisations meet the criteria of intelligent organisation to a larger extent than small and middle-sized enterprises.

Further analysis of the solutions used in Polish enterprises shows that innovative enterprises more often apply the solutions of intelligent organisations than companies which are not innovative. The probability that the innovation company would implement intelligent solutions is about twice higher in comparison with a company which is not innovative, which means that the implementation of the innovation process triggers mechanisms leading to the usage of appropriate solutions applicable to intelligent organizations.

In Poland, intelligent organizations do not have a clear innovative profile yet established as the research in this respect became possible after the period of 2007–2013 which was the time of obtaining resources under the Operational Programme Innovative Economy. Only after this period it was possible to find out what types of innovations Polish companies implemented and how many of them fell under this program. Now, when Operational Programme Intelligent Development 2014–2020 started, it is known that a type of innovation is not a factor differentiating companies in terms of their willingness to implement solutions typical for intelligent organizations. Most often implemented were: process innovations (28%), slightly less more often organizational innovations (24%) and product innovations (21%). The tendency to introduce the solutions appropriate for intelligent organizations to practice of business enterprises increases with the size of company turnover, which is most often positively correlated with its largeness. As far as the sector of business activity of particular enterprise is concerned, intelligent organisations have the biggest share among industrial companies (14%), as well as trade and service companies.

Intelligent organizations are more common among companies with greater turnover. These results are in line with expectations, as the meeting of the criteria of intelligent organizations, require financial investment, which at low scale of operations is not always profitable.

The results indicate then a stronger focus on technological development among intelligent organizations, their better adaptation to the challenges of the knowledge based economy, where the speed of access to knowledge and the possibility of its use is a key factor of competitiveness.

Intelligent organizations in Poland more often use ICT solutions to support management processes in comparison with other organizations. The most commonly introduced is the system for electronic circulation of documents (e-workflow) and databases and data warehouses (83%), as well as Intranet (76%). Other solutions are being used much less frequently—every fourth intelligent organization uses Customer Relationship Management (twice more often than organizations that do not meet criteria for Figure 2. Software solutions as a tool that supports management processes by intelligent organizations and other intelligent organizations) and solutions that support group work, every fifth—the practice of Human Resources Management (having marginal use in remaining organizations) and every sixth—Business Intelligence (three times more often than other organizations) (Figure 1).

Intelligent organizations much more often than other organizations are planning to implement

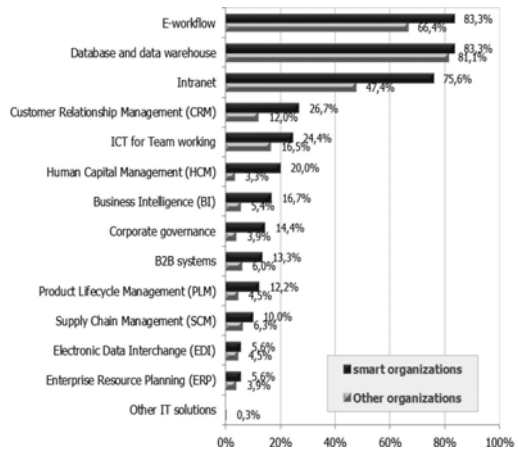


Figure 1. Adopted computer software solutions as tools that support knowledge management processes by intelligent organizations and other organizations.

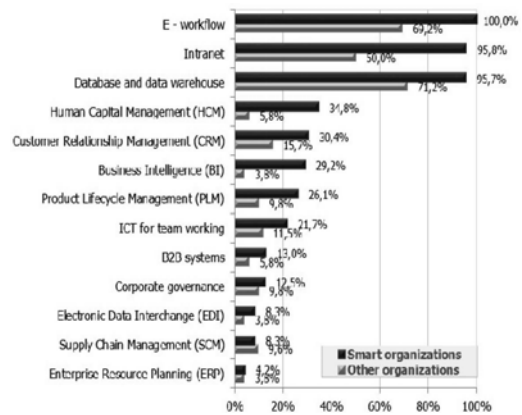


Figure 2. Software solutions as a tool that supports management processes by intelligent organizations and other organizations.

ICT solutions to support management processes. Respondents believe that the vast majority of companies plan to implement electronic document flow (100% intelligent organizations and 69% of other organizations), databases and data warehouses (96% and 75%) and Intranet (96% and 50%). One in three intelligent organization is planning in this period to implement more advanced solutions, namely, human capital management, customer relationship management or business intelligence (Fig. 2).

The last problem that concerns ICT solutions that support the management of intelligent organizations is an assessment of their effectiveness. Enterprises that have used various ICT tools

generally highly evaluated their effectiveness. The few critical comments were focused on low efficiency of databases and data warehouses. Very positively were evaluated the results achieved by the implementation of Supply Chain Management (78%) and Customer Relationship Management (70%). As far as the effectiveness of various ICT tools by intelligent organizations is concerned, but taking into account the size of organisations, it is worth to emphasize that generally ICT tools are assessed as less effective by small businesses than by middle sized and large. This is due to the specific nature of these tools, which do not necessarily have to be effective in organizations with poorly developed organizational structure and not very complicated processes. The respondents representing small companies indeed pointed out that these tools are useful, however, in this group of only a few people indicated very high usability of ICT tools (Kordel, Kornecki, Kowalczyk, Krawczyk, Pylak, 2010).

Finally, Polish companies are still learning how to create the intelligence. This is especially a challenge for companies from the SME sector. In Poland the situation of SME sector companies is changing dynamically as there appear opportunities for the implementation of computerization of companies, in particular in the field of e-business, implementation of ERP systems and dedicated systems, or advanced ICT applications. At the moment no studies that show how these solutions have contributed to the development of intelligent organizations and what is the level of growth of the intelligent enterprises between 2010 and 2015 have been published.

In 2010 MIT Sloan Management Review partnered with the IBM Institute for Business Value conducted a research among nearly 3,000 executives, managers and analysts working across more than 30 industries and involved intelligent organizations of various sizes in more than 100 countries. There were also interviewed academic experts and subject matter experts from a number of industries and disciplines to understand the practical issues facing intelligent organizations today (LaValle, Hopkins, Lesser, Shockley, Kruschwitz, 2010).

As a result, the survey provided following results:

- Intelligent enterprise is focused on the biggest and highest value opportunities against to traditional organizations;
- Intelligent enterprise uses each business opportunity, starting with questions, not data. Traditionally, organizations are tempted to start by gathering all available data before beginning their analysis. Too often, this leads to an all-encompassing focus on data management—collecting, cleansing and converting data that leaves little time, energy or resources to understand its poten-

tial uses. Intelligent organizations should first define the insights and questions needed to meet the big business objective and then identify those pieces of data needed for targets. They can target specific subject areas, and use readily available data in the initial analytic models;

- Intelligent enterprise drives actions and delivers value. New methods and tools to embed information into business processes using cases, ICT analytics solutions, optimization, workflows and simulations are making insights more understandable and actionable;
- Intelligent enterprise develops existing capabilities adding new ones. To do this, it uses sophisticated modelling and visualization tools based on ICT. On the contrary, new tools should supplement earlier ones, or continue to be used side by side, as needed;
- Intelligent enterprise uses an information agenda to do plan for the future. Nowadays, Big Data is getting bigger. Information is coming from interconnected supply chains. Strategic information arrives through unstructured digital channels: social media, smart phone applications and an ever-increasing stream of emerging Internet-based gadgets. The information agenda identifies foundational information practices and tools while aligning IT and business goals through enterprise information plans and financially justified deployment road maps. This agenda helps establish necessary links between those who drive the priorities of the organization by line of business and set the strategy, and those who manage data and information. A comprehensive agenda also enables managers to keep pace with changing business goals. It provides a vision and high-level road map for information that aligns business needs to growth. (Hopkins, Lavalle, Balboni, 2010).

4 ICT DRIVERS EMPOWERING THE INTELLIGENT ENTERPRISE

The literature review and the results of the research conducted by PARP and MIT Sloan Management Review partnered with the IBM Institute for Business Value became the input do the own research conducted in a group of 20 enterprises that the author collaborate with, run to following results.

4.1 *Mobile workforce integration*

In modern organizations access to specific knowledge is critical and mobile connections to operating systems, applications, platforms are important for intelligent enterprises that want to operate efficiently and effectively in fast-paced business environment. Mobile technologies drive technical

innovation to improve networks, ensure employees remain fully integrated with their company and clients wherever they are. Thus, in the intelligent organization its coherency is determined by the intelligence of its network that becomes the organization with wireless tentacles spreading from it to embrace location-aware services—not just for the benefit of tracking, tracing, safety and security, but also as a means to promote the organization itself through marketing and sales initiatives.

4.2 *Smart virtual workplace*

As the approaches to virtualization of IT infrastructure, networks and storage devices continue to mature, infrastructures become software-driven, and IT management more efficient. This efficiency enable services to be provided dynamically according to enterprise, business partners or customer requirements. Smart virtual workplace provides end to end desktop virtualization allowing employees to access applications, data safely over any network from the device of any choice. New trends show that business will increasingly turn to hybrid cloud solutions to enable scalable business processes. Many companies use public clouds for less sensitive applications, but they prefer private clouds for their vital processing tasks with allocation of these tasks as well as data storage for each application being controlled by cloud and edge terminals. Hybrid clouds can quickly scale to a company's needs and services can be paid for as needed. They combine the best of two worlds, offering true benefits to intelligent enterprises aiming to stay ahead in their markets.

4.3 *E-collaboration*

E-collaboration is the de facto standard for business communication, nearly eliminating the need for business travel to meet in person. While knowledge sharing increases, formal and informal groups become collaborative communities that provide coaching and create harmony to reach personal, group, and organizational goals. Intelligent enterprises continue to integrate these into their business processes and reinvent their customer engagement models.

Unified ICT tools allow wide spread teams to work together in real-time, enabling multiple individuals to interact as efficiently and effectively with co-workers, clients, and suppliers.

4.4 *Business flexibility*

The intelligent enterprise must deal with both the complexity of strategy and daily competitive demands. As a result, ICT and support systems must now be highly flexible and resilient in order

to seamlessly communicate and interoperate with other disparate technologies and systems. The cloud architecture enable flexible deployment, therefore the IaaS model becomes very useful. It makes it easier the inter-networking and deploying servers and endpoints from multiple sources.

4.5 *Scalability and customization*

Intelligent enterprises align their IT infrastructure capabilities with business requirements. Modularity of systems, applications allow companies to have only what is needed at present, trimming up-front costs and leaving open the possibility of expanding or incorporating new technologies in the future. With the increase of consolidation, intensive virtualization, the traditional data center environment will transform to the 'hyperscale' data center. It requires a fundamentally different approach than that taken with typical enterprise IT systems. Rather than building 'monolithic' platforms, distributed architecture design is implemented around distributed processing frameworks. That requires software and tools automating node deployment, recover from failure (rerouting of workloads), and other management and monitoring tools.

4.6 *Business continuity*

It is obvious that companies across the globe need to have 24 hours a day access to their data. Data digitalization and rapidity of their processing require more accurate, reliable and sophisticated IT tools converting all data into intelligence for better business outcomes. On the other hand, managers need them to be not complicated in their use. For IT managers there is a challenge. Moreover, for a high level of operational uptime, infrastructure components must be fault tolerant with the ability to recover from complex failures and data storage must be secure. In this case, clustering provides the best solution when insuring uninterrupted workflow on standby systems when failure strikes. This clustering can take the form of software or fault tolerant server solutions which deliver exceptional uptime through dual modular hardware redundancy. These servers will provide continuous availability for all components resulting in optimal data integrity.

4.7 *Converting data into business intelligence*

Advanced ICT solutions enable extracting from huge amounts of data collected from the real cyberspace. Intelligent enterprises are able to manage Big Data projects to drive better business intelligence, product development, and customer service. The important is the fact that they enable to use effectively unstructured data captured from

different systems, mobile devices, social media, log files, emails to perform real-time context analytics. Contextually aware presence allow employees to understand received information, its content to make right decisions in the right time.

Therefore, intelligent enterprises are not only the users of advanced tools based on ICT technologies to optimize business practices, drive workforce engagement and create a competitive edge, but they are also able to leverage and to create value from the data and information generating by ICT solutions.

5 CONCLUSIONS

According to European 2020 strategy for intelligent, sustainable and inclusive growth the EU puts forward three mutually reinforcing priorities. The first is smart growth: developing an economy based on knowledge and innovation. The second, sustainable growth: promoting a more resource efficient, greener and more competitive economy. The third priority is inclusive growth: fostering a high-employment economy delivering social and territorial cohesion. The document includes targets for the whole European community to achieve. New goals and priorities concern many fields of science, economy, business management and education, but above all they focus on the sector of research and development (OECD, 2004; European Commission, 2010). Therefore so important is the use of information technologies that are shaping almost every aspect of modern business. In this situation, the key factor are the resources and the effectiveness of the use of ICT solutions combined with the management of flow of information, knowledge and financial resources.

The research conducted by Polish Agency for Enterprise Development showed that small and medium-sized enterprises in Poland have still many challenges to overcome. They must change a lot to deserve to be called intelligent organizations. In comparison with other EU countries, Polish companies use popular ICT tools, commonly available, as they do not have enough funds for investment in comparison to their competitors. The concept of the intelligent organization is relatively new in the context of the modern enterprise management style, which was created in response to the ever-growing competition, and high pace of technology development and rapidly changing business conditions that require skilful combining knowledge that the organization has with and what the technological achievements offer.

Research conducted by MIT Sloan Management Review partnered with the IBM Institute for Business Value showed that the intelligent

enterprise must be able to manage data and information effectively using appropriate analytical tools based on ICT. The way in which ICT tools are exploited depends on the knowledge and skills of managers and analysts who in this case play a significant role.

Research conducted by the author correlate with the above findings. They represent the first approach to the problem of an intelligent organization that in doing business uses a variety of ICT—based solutions. The author made an attempt to identify which of the ICT drivers have the biggest influence on shaping an intelligent organization. The advanced research will be continued in the nearest future.

REFERENCES

- Christensen, C. M., *Making strategy: Learning by doing*, Harvard Business Review 1997, no. 4, 141–156.
- Davenport T.H., Leibold M., Voelpel S., *Strategic Management in the Innovation Economy*, Wiley, Erlnegn, Germany 2006.
- European Commission, Europe 2020: *Strategia na rzecz inteligentnego i zrównoważonego rozwoju sprzyjającego włączeniu społecznemu*, Brussels, 3.3.2010, KOM (2010).
- Hamel, G., Prahalad C. K., *Competing for the Future*, Harvard Business School Press, Harvard 1994.
- Hopkins M. S., Lavallo S., Balboni F., *10 Insights: A First Look at the New Intelligent Enterprise Survey on Winning with Data*, MIT Sloan Management Review 52, no. 1, 2010.
- Kordel P., Kornecki J., Kowalczyk A., Krawczyk K., Pylak P., Wiktorowicz J., *Inteligentne organizacje – zarządzanie wiedzą i kompetencjami pracowników*, Polish Agency for Enterprise and Development, Warszawa, 2010.
- LaValle S., Hopkins M.S., Lesser E., Shockley R., Kruschwitz N., *Findings of the New Intelligent Enterprise Study*, Big Idea: Data & Analytics, October 24, 2010.
- Muryjas P., Wawer M., *Business intelligence as a support in human resources strategies realization in contemporary organizations*, Економіка та управління підприємствами, 2014, p. 183–189.
- Nonaka, I., Takeuchi, H., *The Knowledge Creating Company*, Oxford University Press, Oxford 1995.
- Organisation for Economic Co-operation and Development—Centre for Educational Research and Innovation, *Innovation on the Knowledge Economy—Implications for Education and Learning*, OECD Publishing House, Paryż 2004, pp. 14–15.
- Schwabinger M., *Intelligent Organizations: Powerful Models for Systemic Management*, Springer-Verlag, Berlin Heidelberg, Germany 2009.
- Thannhuber M. J., *The intelligent enterprise: theoretical concepts and practical implications*, Physica-Verlag, A Springer Company, New York 2005, p. 72.
- Wilensky, M.L., *Organisational Intelligence*, Basic Books, London 1967.

The presence of sustainable entrepreneurship in Polish companies based on the selected examples

Paula Bajdor

Faculty of Management, Częstochowa University of Technology, Częstochowa, Poland

ABSTRACT: Entrepreneurship can be seen as a major factor influencing both, economic and social development. The desire of man to continuous development by investing new products and services, and then taking implementation actions is the essence of entrepreneurship. In the popular opinion, the entrepreneurship means taking the challenges, actions and activity in the professional and private life. Due to the fact that entrepreneurship is manifested in almost every aspect of life, it has been known for several years, that it has a significant contribution in implementing the concept of sustainable development, which resulted in the emergence of the term “sustainable entrepreneurship”. The aim of the article is to present the essence of sustainable entrepreneurship and its presence in companies operating in the Polish market. The study has resulted from the literature review on entrepreneurship and then the selected examples of companies, operating in the frame of sustainable entrepreneurship has been presented. The analysis has shown that in Poland there are companies conducting its activity according to basic principles of the discussed subject but not so many so far.

1 INTRODUCTION

Entrepreneurship, next to flexibility, setting goals and striving to achieve them, is one of the characteristics of human resources. The market conditions, in which the entrepreneurs have a freedom to act, fully serve to entrepreneurship's development. Entrepreneurship is also an important factor in achieving their success, which is usually measured by the height of the gained profit, it is also a force that integrates a variety of resources, which are necessary to start an economic activity (Piasecki 2007). Thanks to the entrepreneurship, new enterprises have been created and thus affect the creation of new job positions, so the tangible and intangible values are multiplied (Kadłubek, Lis 2013). Above all, entrepreneurship contributes to the competitiveness of the organization not only in the closest environment but also on a global scale.

Entrepreneurship is an equivocal term and attempts to define it, have been taken by many authors dealing with this issue. However, J.A. Schumpeter is considered as the creator of the theory of entrepreneurship, who defined it as a process of creative destruction, leading to new combinations in the sphere of production creation, being the germ of entrepreneurial activities (Samuelson, Nordhaus 2006). And entrepreneur is responsible for creating new economic ways, which have brought solutions better than the previous. The entrepreneur is usually defined as one who organizes, manages

and assumes the risk of a business or enterprise (Graham 2010).

According to this, the essence of entrepreneurship lies in breaking with routine and dismantling existing structures. Another definition of entrepreneurship describes it as willingness and ability to make and solve emerging problems in a creative way, and the ability to adapt to changes in the environment (Strużyński 2033). The stimulator of dynamic development of entrepreneurship and the creation of small businesses willing to take the risk of economic self-employed is the situation on the labor market.

Entrepreneurship is being also analyzed in terms of the life cycle of an organization. In almost all cases, it occurs in the first stage of the life cycle—as a step in the formation of the organization and business initiation (Chodyński 2009). It can be assumed that to a certain extent, entrepreneurship determines the formation of organization. Entrepreneurship also appears in subsequent life cycles but usually with stabilizing or ordering functions. In the stronger degree, entrepreneurship manifests itself in the last cycle of organizational life, when the situation in which it is necessary to conduct a kind of “renewal” or restructuring of the organization most common occurs.

In the literature on entrepreneurship, the business strategy term distinguishes, whose main role is to use opportunities during the strategic development of the organization (Yu, Huarng

2013). In this regard, entrepreneurship is based on the interaction of development opportunities, strategic factors and the competence of the organization. These last ones are considered as the skills, necessary for full coordination of resources, used to achieve the intended goals of the organization (Starostka-Patyk 2014).

In addition, in the literature on entrepreneurship, we can find different business models presenting different types of entrepreneurship including sustainable entrepreneurship.

2 THE CONCEPT OF SUSTAINABLE ENTREPRENEURSHIP

The sustainable entrepreneurship concept has gained importance over recent years. The concept of sustainable entrepreneurship was derived from sustainable development concept, which can be defined as *the continuing commitment by business to behave ethically and contribute to economic development while improving the quality of life of the workforce, their families, local communities, the society and the world at large as well as future generations* (Gagnon 2012). It links sustainable development with business activities. And this relationship has been dealt with through various ways of thinking, and resulted in the appearing of new types of entrepreneurs, called as ecopreneur or social entrepreneur (Wu, Huarng 2015). What

is the difference between them, is that ecopreneur pursues environmental opportunities from an economic point of view, while the social entrepreneur—is characterized by a broader social perspective (Kot, Brzeziński 2015). Recently a new term “sustainopreneurship” has been shown, it is a portmanteau of sustainability and entrepreneurship (Abrahamson 2009). And it described the business organization in a way to solve problems related to social and environmental sustainability. It is simply “business with a cause” – where main problems of current worlds are turned into business opportunities by deployment of sustainable innovations (Brzozowska, Dacko, Kalinichenko 2015).

Sustainable entrepreneurship is called by some researchers as a “sustainability-driven entrepreneurship” or “sustainable-minded entrepreneurship”. But is some studies an “environmental entrepreneurship” can be found (Okwiet 2013). So many various terms cause that, at present, there is no universally accepted definition on sustainable entrepreneurship and in the literature many definitions proposed by many researchers can be found. Definition given by Majid and Koe, describes sustainable entrepreneurship as *an entrepreneurial process to exploit the opportunities in an innovative manner for economic gains, society equity, environmental quality and cultural preservation on an equal footing* (Majid, Koe 2012). Another definition given by Dean and McMullen says: *sustainable entrepreneurship is the process of discovering, evaluating, and exploiting eco-*

Table 1. The main differences between these concepts.

	Ecopreneurship	Social entrepreneurship	Institutional entrepreneurship	Sustainable entrepreneurship
Main motivation	Solving the environmental problems together with economic value creation	Solving social problems and create a value for population	Contribute to changing regulatory, social and market institutions	Contribution through solving social and environmental problems together with a successful business activity
The main purpose	Raising money by solving the environmental problems	Achieve the social goals and secure funding	Changing the institutions	Influencing the SD through entrepreneurial corporate activities
The role of economic goals	Ends	Means	Ends or means	Ends and means
The role of non-market goals	The core element are the environmental issues	Social goals	Changing the institutions	Ends integration to sustainable development
Organisational development challenge	From focus on environmental issues to integrated economic ones	From focus on society issues to integrated economic ones	From changing the institutions to integrating sustainability	From small contribution to large contribution to sustainable development

Source: Own work based on Di Maio 2011.

conomic opportunities that are present in market failures which detract from sustainability, including those that are environmentally relevant (Dean, McMullen 2007). S. Graham describes it as *the process of sustaining a level of entrepreneurial development as to create a paradigm shift in economic activity such that national GDP, job growth, capital investment, technology advancement, and quality of life is unmatched, unsurpassed and unequalled. And the human population should strive through local, state and national efforts to seek to establish an economic mentality that is strategically focused on entrepreneurship and authentic organic economic growth at the community level* (Graham 2010). While Shaltegger and Wagner perceive sustainable entrepreneurship as *in essence the realization of sustainability innovations aimed at the mass market and providing benefit to the larger part of society. By realizing such a (radical) sustainability innovations sustainable entrepreneurs often address the unmet demand of a larger group of stakeholders* (Shaltegger, Wagner 2011).

At present sustainable entrepreneurship has been perceived as an overarching way of looking at the entrepreneurs' contribution to ecological, economic and social aspects. Thus, sustainable entrepreneurship today, is perceived as a way of generating competitive advantage by identifying sustainability as a new business opportunities. Which may result in new, more sustainable productions methods, business organization and products (Schlange 2006). What is characteristics for sustainable entrepreneurship is that this concept is not about meeting environmental regulations only, but it rather takes advantage of the increasing need to more sustainable production resulting in more sustainable products and services.

3 THE SUSTAINABLE ENTREPRENEURSHIP MODEL AND PRINCIPLES

Due to the fact that sustainable entrepreneurship is a very broad issue, the authors dealing with this subject, in addition to the definitions, distinguish the main principles, which fulfillment leads to this concept implementation. Based on the literature sources it is possible to identify the following rules (Di Maio 2011) (Pascual, Klink, Grisaes 2011):

1. The use of resources in an economical way, and their excessive use should not be the goal of life.
2. The accumulation of goods should not be the purpose of life as well, as an economic good does not mean possession of large quantities of goods.

3. Reducing the level of consumption and then—waste.
4. Long-term goals should be just as important as short-term—taken actions should be considered also in the wider perspective and not just in short-term benefits, as sustainable entrepreneurship is a journey not a goal.
5. Knowledge has the greater value than financial means.
6. The ethic and justice laws should be kept, because striving to obtain benefits at any cost is simply “uncool”.
7. Consider environmental, social and financial value while formulating new ways of activity.
8. Pay attention on sustainable innovations, as they allow to develop unforeseen opportunities.
9. People play an important role and are the grater source of opportunity, the only limit is a science.
10. It is important to understand the overall context of the problem and identify the opportunities.
11. The entrepreneurs should create a value for the people around.
12. Use the power of information—obtain it, use it, share it, transform it, mix it and reflect upon it.
13. The entrepreneurs should not be limited by their mind.

The principles presented above show that sustainable entrepreneurship does not only refer to the entrepreneur, but also to a certain community or traders. Taking action according to some of them, causing already that these companies can be described as “sustainable”. Next to these main principles, in the literature a model proposed by W. Young and F. Tilley, called “The sustainable entrepreneurship model” can be also found (Tilley, Young 2007).

In this model we can identify environmental entrepreneurship, economic entrepreneurship and (social entrepreneurship), it also presents an additional element called sustainable entrepreneurship, which with all, earlier mentioned, elements establishes relations cover the following (Tilley, Young 2008):

1. The relationship between entrepreneurship and balanced economic entrepreneurship includes:
 - a. economic capital, the distribution of which takes into account both current and future generations,
 - b. Capital intergenerational, which refers to ensure the existence of future generations, and its distribution takes into account, during taking the actions and decisions of the organization,

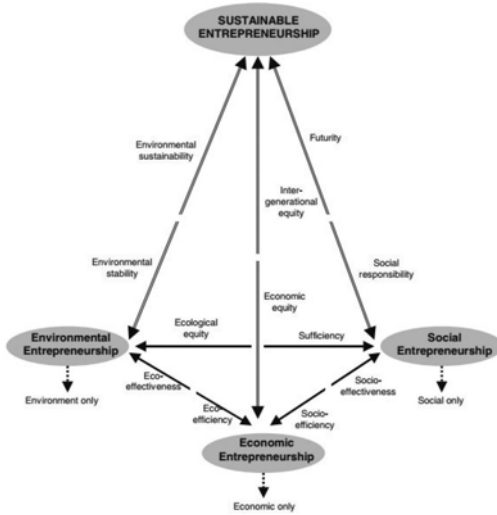


Figure 1. The sustainable entrepreneurship model. Source: Own work based on (Tilley, Young 2008).

4 THE SELECTED EXAMPLES OF POLISH COMPANIES OPERATING ACCORDING TO SUSTAINABLE ENTREPRENEURSHIP PRINCIPLES

4.1 First example—a hotel

The hotel which is the subject of this case, has been operated since 2001, and was supposed to be a small hotel, characterized by the relaxing and “at home” atmosphere. From the beginning, the hotel was very equipped and its staff consists of highly qualified and experienced people, it makes that for a few years, the hotel has had a good reputation among visitors to the Czestochowa city. At present it has 31 rooms which can host more than 60 people, and for several years, the hotel has a restaurant, which filled the hotel offer on the part of the catering. Initially, the hotel owners have focused primarily on reaching as many people, by taking action, based on hotel advertising in different parts of country and beyond its borders. However, for several years they undertake the activities that fit not only in the principles of sustainable development but also in the framework of sustainable entrepreneurship. Mainly these actions consist of:

2. The relationship between environmental entrepreneurship and sustainable entrepreneurship includes:
 - a. environment stabilization, which is subject to appropriate forces that stabilize it, as well as lead to the restoration of some ecosystem services, which, for example, slow down the processes of climate change,
 - b. environmental sustainability, which refers to its long-term stability, this aspect is also taken into account in taking the actions or decisions by organizations
3. The relationship between social entrepreneurship and sustainable entrepreneurship includes:
 - a. Social Responsibility, whereby companies and individuals take responsibility for direct and indirect, positive and negative effects of actions affecting the present generation,
 - b. When taking actions and decisions by the organization, the future prosperity of the next generation, is also taken into account.

- Monitoring the water and electricity consumption by using special counters, as well as seeking opportunities to reduce their consumption by installing, in the case of energy—saving light bulbs, and in the case of water—battery-operated photocells,
- Reuse of water, eg for watering plants and lawns,
- Customize the type of energy-saving light sources to the requirements of individual rooms,
- Four years ago a thorough insulation of the building has been conducted—with the replacement of windows and the use of better insulation, the building loses much less heat than before,
- Segregation of waste introduction, including the purchased suitable containers for different kinds of waste,
- The grocery products made from environmentally clean sources of supply, located in the immediate area, now such purchases represent approx. 20% of all purchased food products,
- All employees of the hotel have a contract of employment on which contributions are paid, and paid salaries correspond to those which are specified in the contracts,
- There are valid safety regulations and the newly recruited employees are directed to the training of their field, before they take their professional obligations,
- The hotel takes an active part in the activities undertaken for the development of local culture,

The presented model indicates that sustainable entrepreneurship consists of 12 elements, but in the course of conducting more detailed research on sustainable entrepreneurship, there may also be another, essential elements. This model, however, can be a useful tool, used in examining whether the organization can assign a sustainable feature or it still should take actions in this direction.

usually in the form of patronage or in the form of sponsorship support. In the case of a larger cultural events, the hotel provides free accommodation for the guests, for example. Actors or artists

- Its 1% tax transfers to the city

As seen from the above, most of the actions taken by the hotel are in line with the ecological aspect of sustainable entrepreneurship, but the remaining aspects were also taken into account. They are reflected attention to employees and commitment to the city.

4.2 *Second example—a production company*

Company B is a printing company and has operated in the local market since more than 10 years, and during that time has gained a wide range of customers. The nature of the company has been identified not only by the posters, folders or flyers, but also by a large amount of waste paper, empty containers of paint and large amounts of toner cartridges for printers. A few years ago, the top management of the company has decided to implement a solution, that not only would reduce the amount of waste produced, but contributed to significant savings in expenditure associated with waste paper or toner. As a result of the analysis of the potential areas for improvement, the company has implemented the following solutions:

- Separation of printers and printer working professional. In addition, each employee has been assigned a code by which it becomes possible to identify, who and how many printed within a week or month. Any deviation from the average output for the period is explained in detail.
- All waste in paper form are stored in appropriate containers, and collected by the company recycler, in return the company gets recycled paper at greatly reduced prices,
- Toner used in printers cartridges are not original but are so called fillers, which can be repeatedly used. Currently, the level of toner print such does not differ from the level of quality prints of original toner cartridges, and this type of printing allows for significant cost savings for the purchase of toner,
- Over a year ago, the company bought strips with circuit breakers, on each strip for a single job. After the work, employees are required to exclude not only computers but also all the slots, which makes the whole electronic equipment is actually turned off, and consume no power,
- On computers has also installed a program, which add to every sent email a footer with information, that informed about that when there is no need to print the document, do not print it,

- In addition, employees of the company were required to bring their own dishes and cutlery as well as introduced a ban on carrying plastic bags into the company,
- for the local community, the company takes an active parts in all local events such as festivals, fairs and local festines. Almost always it provides flyers about the events for free and sometime participate in organization costs,
- unfortunately it was impossible to identify any action in order to fulfil company's employee's needs—only few of them have a regular contract as the most of them work on service contract. The company does not send its employee on any additional workshops and courses. The only thing, which can be regarded in term of "social" entrepreneurship is fully established work and safety policy. Ane in case when the employee needs to take an additional day off—the company allow for it, but employee's salary has been cut off due to day off.

Again, from the above, it can be concluded that, again the environmental aspect is the most fulfilled one. But the social aspect is almost not visible.

4.3 *Third example—another production company*

Company C is a manufacturer and distributor of pushchair and kids accessories. Its headquarters is in Czestochowa, and has more than 20 years of experience in manufacturing strollers and baby products. This company, from all the mentioned, can be described as the one, who fully meets the principles of sustainable entrepreneurship. What is visible through the following implemented solutions:

- The entire lighting companies have been exchanged into LED bulbs, that give light similar to natural while reducing power consumption,
- Also in the toilets and kitchenettes, are fitted batteries acting on the photocell, the water begins to flow at the time of planting hands on the tap and stops flowing immediately after their withdrawal,
- All waste, both from production and from the administrative offices, are selected and collected by the companies involved in recycling. Sorting of waste and its reporting to the relevant companies, resulted that the company is not burdened charge for garbage collection,
- Production hall and warehouse are equipped with adequate ventilation and doors, that are very well insulated and do not leak heat to the outside,
- Since last year, in the sample, several employees of the company working for three days a week at home to the headquarters of the company come only on Mondays and Tuesdays. This solution also contributed to higher energy and water, not to mention the same employees who really this mode of work,

- In addition, the company is trying to implement a policy of car-sharing, as a significant part of the workforce commutes to workplaces, established plaque on which employees could enter the place of departure with time. It turned out that there were a lot of employees on the same route, each in his car. At this point, formed several groups of workers who commute to the company one car, sharing fuel costs among themselves,
- Also, the company takes an active part in the event called “The noble box project”—the project was established in 2001 in order to provide aid to struggling families during Christmas holidays (www.szlachetna.paczka.pl). Every year, the company sends a dozen of packages—with puschchair and kids accessories to selected families.
- Occasionally the company becomes a sponsor of selected cultural and sport events organized in the city.

It seems from the above, that the this example fully presents the essence of sustainable entrepreneurship identified in the third company. It cares not only about the environment, but also takes account the social aspects. Together with these two, the economic aspects can be perceived by higher and higher profits achieved by the company.

5 CONCLUSION

The article has discussed the essence of sustainable entrepreneurship according to activities taken by polish companies. Threat with regard to sustainable entrepreneurship is that it can be treated as a synonym for sustainable development. Because very often the actions to follow up processes of sustainable entrepreneurship can be regarded as a practical embodiment of the principles of sustainable development. However, while the concept of sustainable development has a much more balanced overall dimension, sustainable entrepreneurship is mainly reflected in practical terms, in the form of concrete actions enrolling in the frame. Presented in this article companies fit into the concept of sustainable entrepreneurship, while taking steps to improve not only the environment, but also improve the position of their employees and immediate social environment. In contrast, due to the fact that these are private companies, it is clear that their activity is focused mainly on achieving profit.

REFERENCES

Abrahamsson, A. 2006. *Sustainablepreneurship—Business with a Cause. in Science for Sustainable*

- Development—Starting Points and Critical Reflections*, Swedish Society for Sustainable Development, Uppsala, pp. 21–30.
- Brzozowska A., Dacko M., Kalinichenko A. 2015. *Sources and Determinants of Enterprises’ Innovativeness*, *Aktual’ni Problemi Ekonomiki* vol. 9/171, pp. 182–188.
- Chodyński A. 2009. *Przedsiębiorczość i innowacyjność a kompetencje—aspekty strategiczne*, *Zeszyty Naukowe Wyższej Szkoły Humanitas*, Sosnowiec.
- Dean T.J., McMullen J.S. 2007. *Toward a theory of sustainable entrepreneurship: reducing environmental degradation through entrepreneurial action*. *Journal of Business Venturing*, 22 (1), pp. 50–76.
- Di Maio P. 2011. *Sustainable Innovation*, ISTCS, Edinburg.
- Gagnon, M.A. 2012. *Sustainable Minded Entrepreneurs: Developing and Testing a Value-based Framework*. *Journal of Strategic Innovation and Sustainability*, 8(1), 9–25.
- Grabara J., Bajdor P., Mihaescu L. 2015. *Steps of Sustainable Development Implementation Into Enterprise Activities*, *Management of Sustainable Development*, vol. 7/1, pp. 45–49.
- Graham S. 2010. *What is Sustainable Entrepreneurship*, Ezine Articles, United States.
- Kadłubek M., Lis T. 2013. *Innowacyjność organizacji w aspekcie logistycznej obsługi klienta*, *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, nr 310, pp. 339–347.
- Kot S., Brzeziński S. 2015. *Market Orientation Factors in Sustainable Development and Corporate Social Responsibility*, *Asian Journal of Applied Sciences*, vol. 8/2, pp. 101–112.
- Majid, I.A., & Koe, W.L. 2012. *Sustainable Entrepreneurship (SE): A Revised Model Based on Triple Bottom Line (TBL)*. *International Journal of Academic Research in Business and Social Sciences*, 2(6), 293–310.
- Majid, I.A., Kamaludin, M.H., Saad, M.S.M., & Aziz, N.A. 2012. *Sustainability-driven Entrepreneurship: The Mediating Effect of Opportunity based Management Structure on the Relationship between Entrepreneurial Orientation and Environmental Sustainability Management of SMEs: A Conceptual Framework*. *European Journal of Business and Management*, 4(13), 148–155.
- Okwiet B. 2013. *Entrepreneurship in small and medium enterprises sector—development barriers and opportunities*, *Polish Journal of Management Studies*, vol. 7/2013, pp. 37–47.
- Pascual O., Klink A., Grisales J.A.R. 2011. *Create impact! Handbook for sustainable entrepreneurship*. *Enviu—innovators in sustainability*, Rotterdam, pp. 1–54.
- Ptak A., Sroka M. 2014. *Entrepreneurial Activity in the EU Member States*, *International Conference on Horizontal Approaches in Education and Culture in the Context of European Macro-Strategies*, pp. 63–69.
- Schaltegger S., Wagner M. 2011. *Sustainable entrepreneurship and sustainability innovation: categories and interactions*. *Business Strategy and the Environment*, 20, pp. 222–237.
- Schaltegger, S., & Wagner, T. 2008. *Types of Sustainable Entrepreneurship and Conditions for Sustainability*

- Innovation: From the Administration of a Technical Challenge to the Management of an Entrepreneurial Opportunity.* In R. Wüstenhagen, J. Hamschmidt, S. Sharma, & M. Starik (Eds.) *Sustainable Innovation and Entrepreneurship* (pp. 27–48). Glos: Edward-Elgar.
- Schlange, L.E. 2006. *What Drives Sustainable Entrepreneurs?* 3rd Applied Business and Entrepreneurship Association International (ABEAI) Majid, Koe Conference, Kona, Hawaii.
- Starostka-Patyk M. 2014. *The General Idea of Environmental Management Development*, I. International Scientific Conference Green Energy—Environment—Sustainable Development, Presov, pp. 87–93.
- Strużyński M. 2003. *Przedsiębiorstwo a rynek*, PWE, Warszawa.
- Ślusarczyk B., Broniszewska A. 2014. *Entrepreneurship of Women in Poland and the EU—Quantitative Analysis*, Polish Journal of Management Studies, vol. 9/2014, pp. 217–224. www.szlachetna.paczka.pl.
- Tilley F., Young W. 2007. *Can Business Move Beyond Efficiency? The Shift toward Effectiveness and Equity in the Corporate Sustainability Debate*, Business Strategy and the Environment, nr 15, p. 410.
- Wu C.W., Huarng K.H. 2015. *Global entrepreneurship and innovation in management*, Journal of Business Research, vol. 68/4/, p. 743.
- Yu T.H.K., Huarng K.H. 2013. *Entrepreneurial firms' health creation via forecasting*. The Service Industries Journal, vol. 33, p. 835.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Raising demand for implementation of systemic logistics management activities in agribusiness

Anna Brzozowska & Katarzyna Szymczyk

Faculty of Management, Częstochowa University of Technology, Częstochowa, Poland

ABSTRACT: The aim of the study is to analyze the problems and barriers related to criteria of logistics management in agribusiness. The study has resulted from the literature review on logistics management system and indicates that the logistics management system is an issue not yet fully explored and undertaken by theorists. This signifies that policy makers as well as farm and rural areas managers may have a problem with correct execution of the rules and solutions necessary in order to achieve the anticipated outcomes of their actions. The field of logistics management, so often ignored, does not give the opportunity to obtain the expected synergies and added value in the flow of goods in the supply chain in agribusiness.

1 INTRODUCTION

Modern logistics defined as “new day logistics” (Sople 2007) puts mainly an emphasis on one of its basic aspects i.e. a systemic approach (von der Gracht 2008). This term points towards different and various components and activities included in the entire supply chain (Ciesielski 2008) and analyzes them as a whole, particularly in the form of a system.

Implementation of logistics specific tasks—planning, organizing, managing and controlling the flow of goods and people takes place between the elements of the logistics chain. The system approach (Neider 2006) does not treat them individually as independent components, but as interrelated and correlated parts of dynamic system. Such system operates on the basis of three fundamental principles attributable to the content of general theory of systems (Fig. 1).

In terms of the system all its components are linked together in terms of network of relationships, where changes made to one component influence the functioning of others and thus determine the efficiency of logistics operations done within the entire system. Therefore, it may simplify finding causes of irregularities in the sphere of unplanned expansion of stores or prolongation of the term of the contract in the logistic system, not only in the part of the system where irregularities occur, but in all other parts of the system (Blaik 2001).

Effective integration of system components implies multiple benefits. Thanks to that we obtain the interaction of all components increasing the efficiency of the system and by this means we drive back the individual actions of individual elements

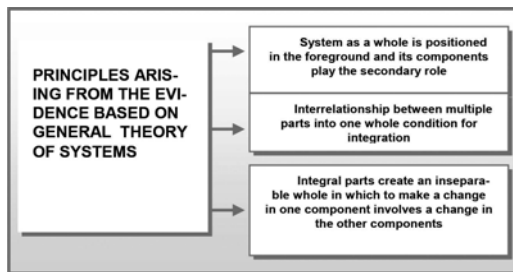


Figure 1. The fundamental principles arising from the evidence based on general theory of systems.

to the background in agribusiness and gain synergies that integrate with the objectives of the Europe 2020 strategy.

Therefore, the rising demand for transpiration of logistics management system activities requires a holistic reasoning, thinking in terms of system and comprehensive way of examining issues, requires a holistic reasoning, thinking in terms of system and comprehensive way of examining issues hence explaining a whole is not just explaining of its elements, but effects of synergies are necessary to transpire in order to give you an idea about the relationship between these elements.

2 CONCEPTUAL, INTERPERSONAL AND TECHNICAL SYSTEM APPROACH TO MANAGEMENT SYSTEM LOGISTICS

The system approach to logistics processes is based on correlation between particular areas (Sołtysik

2003). These areas, already in the planning process, must generate one piece since in future the integrated logistics activities will be anchored in the competent management of processes in agribusiness.

The logistics system consists of various elements forming functional units which enable performance in a amount of one enterprise (at work, the enterprise is acknowledged to agricultural farm). Nevertheless, the enterprise itself can exist as a component of a larger system, which affects the diversity of management's logistics system.

Therefore independently and permanently, logistics system is related to multifactorial management describing standards, strategies of actions, transport activities in different environments (Dlugosz & Zimniewicz 2009) and warehousing, allowing for the cost of ongoing projects (Foltynowicz, Jasiczak & Cone 2008) which should be supported by scientific expertise founded on strategic and logistics management.

Hence, creative approach to economic demand for implementation of activities of systemic logistics management in agribusiness should include all processes related to production and distribution and also identify four elements constituting one unit in the logistics chain, i.e.: procurement logistics, production logistics, distribution logistics, utilization logistics (Fig. 2). If the mentioned above forms of management will not be applied, agricultural farms will not be able to function in the market economy in efficient way.

The range of the system (number and type of institutions in the system) determines the division into three main groups: operating in micro scale, meta- or macro scale (Nowicka-Skowron 2000).

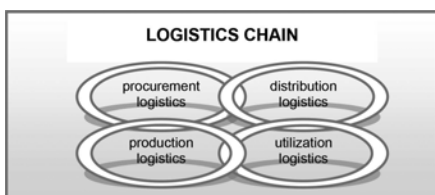


Figure 2. The logistics chain.

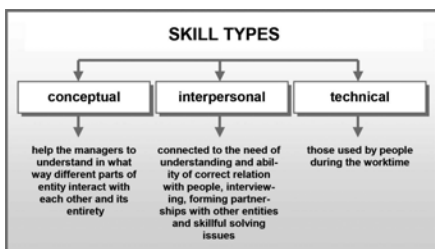


Figure 3. Skill types acquired by individuals implementing management process.

Micro-logistics system can be determined within one company. It is the identical logistics system of a single sub-economic entity functioning in the market and referring to the logistics activities of all kinds of enterprise, i.e.: production, service, or trade (Grzybowska 2009). The system consists of elements in the form of functional units of the company responsible for the course of the business processes, such as supply, production processes, cargo storage, or distribution. In this case, it becomes a part of a logistics chain in meta- or macro-logistics system.

Equally important are the skills acquired by individuals performing management process. Researchers identify three basic types: conceptual, interpersonal and technical (Fig. 3).

Nonetheless, it is difficult to classify various single types to the basic types because of the one possible to occur (Ghillyer 2009) and because, besides agents of coherent management, the logistics process would be poor in any strategy and its nature would become worthless. Such an action would generate unnecessary costs and introduce an unwanted chaos in the performance of the entity. Due to the symbiotic relationship between management and the logistics more and more managers governing the companies appreciate not only the commercial advantages of this correlation, but also have contributed to the development of logistics, the one mentioned above, in many aspects of everyday life (Bruska 2012).

The untroubled functioning of enterprises requires a systemic approach to logistics management. In a market economy, there is a must of building a database and identifying key principles, i.e.: rejecting the idea of eternity and perfection and creating new rules of behaviour and approach that will provide an adequate answer to a globally transforming reality, and thus, will allow effective functioning of the agricultural company (Nogalski 2011).

3 REASONS FOR SUPPORTING LOGISTICS MANAGEMENT AS SUBSYSTEM OF ORGANIZATION

Logistics support is understood as a reflection of the nature of the logistics functions that accompany each systems and socio-economic processes (Chaberek 2002).

In the literature study we find this logistics support system as a coordinated, deliberately organized subsystem of organization that supports its basic processes within the coherence of all activities related to the necessary flow of resources (Fig. 4) (Chaberek 2002; Szymczak, 2006).

On the micro scale there is a demand for logistics support from the production system (Bruska 2012). The effectiveness of logistics support is measured

by the correct amount of appropriate resources for specific primary process at the right time and place. If these objectives are not gained, the effectiveness of logistics operations is not achieved at the expected level (Chaberek & G. 2009).

The first zone of logistics support for agribusiness focuses itself on improving the workflow of processes and their optimization in existing conditions. Another zone moves the idea of optimization of logistics support to the relationship with the fundamental processes taking place in rural areas. They are gradually identified in a partial and descriptive way (Markowski 2006, A. Bruska 2012) which enables improving the core processes together with creating new concepts of logistics management (more about: Christopher, 2005; Stevens 1989), especially, the methods and tools used in improving the supply chains. The essence of supply chain (Christopher, 2005; Kotler, 2005; Waters 2009; Hu-gos, 2003; M. Goldratt 1984), in its assumptions, carries many challenges that lead to a rapid increase of the meaning of the supply chain management (Harrison & van Hoek in 2010; Ketikidis Koh, Dimitriadis Gunasekaran & Kehajova 2008; Manganese, Lalwani & Butcher in 2008; Mikus 2003; Pereira, 2009). The identification of the management in the supply chain (more about: Rutkowski 2004; Hugos 2003; Blaik 2010; Larson & Halldorson 2004; Prockl 2007; Shah in 2009; Altekar, 2005; Bozarth & Handfield 2007; Kearney, 1978) comes to: planning, organizing, managing and controlling (Daft, Kendrick & Vershinin 2010) during the following stages: planning and management of supplies, raw material sourcing, planning the production of raw materials, storage

and delivery to the final consumer. And finally, the third zone corresponds to the relationship enabling the development of logistics support systems with the external services that occur in this process and also, or even first of all, include the activities of public administration and local government.

Taking into consideration the nature and potentiality created by the implementation of the idea of logistics support for the expanding demand for incorporation of activities of systemic approach to logistics management, rural areas build the arena due to its dominant position in the spatial layout.

Thus, the logistics activities reflect the process of fulfilling the demand. Therefore the deeper analysis of the flow system in agribusiness is needed, the system which shows a number of possibilities for optimization due to the search for synergies of logistics management activities which may lead to identification of the comprehensive monitoring system of logistics support and improvement of the flows value in rural areas (Bruska 2012).

In real terms, a comprehensive mapping and the systemic analysis of the logistics management in agribusiness causes the identification of business models. (The business model can be regarded as developed, contemporary, differing from traditional and classic, form of the organizational model of the enterprise and introduce as a structured idea of desired directions of development of the company (its strategy) and conditions of this process (Nogalski 2011) in these areas being included in the type of competitive advantage, especially existing skills referring to the strategy content (Obłój 2002; Brzóska 2007) useful both for the agribusiness sector, as well as agri-logistics (Bruska 2012). The solutions optimizing the functioning of the logistics support in rural areas do not exclude certain decisions that basically answer to questions whether the absence of competitiveness

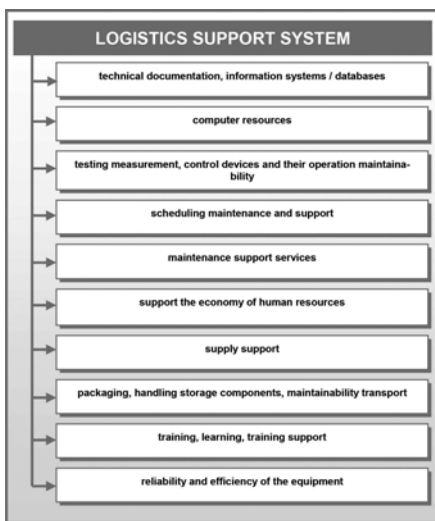


Figure 4. Logistics support system.

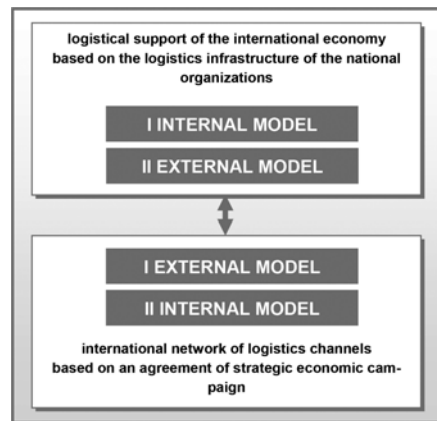


Figure 5. Modes of development of logistics support.

(Competitiveness as a microeconomic category is seen in the relationship as the managing entity, its potentiality, capabilities or skills. More about: Głód 2007) in the agribusiness sector, dependent on various forms of sectoral support, will allow (agri)logistics to develop effectively (Głód 2007).

Through the adjustment of logistics support and its system on a larger scale of activities in agribusiness, the certain enumeration of basic tasks of logistics support process should be mentioned, the process that accomplishes the needs at the community level (Bruska 2012). It includes tasks resulting from fulfilling needs in terms of individual consumption of households, access to services of collective consumption and processes including needs carried out by local producers. Concentration of logistics support in the agricultural enterprises and rural areas is illustrated in Fig. 6.

However, It is not a comprehensive range of tasks the logistics support system faces in rural areas. The diversity of reasons and the forms of flows in rural areas can group them as shown in Figure 7.

According to the above statement, one may come to conclusion that: “The logistics support along with changes in the economic environment is given upon the gradual transformations which allows for distinguishing four models of development. The first two—internal models—carry out logistics support of international economy with the help of the logistics infrastructure of specific national organizations. Another two—external models- are based on strategic agreements between economic companies of different countries and create international network of logistics channels. The important significance belongs to support platforms of the fourth generation which determines in its basics the “transparency” of information in the logistics channel, as well as the creation of virtual organizations in the context of the supply chain. Consequently, the need for the size of working capital in the analyzed logistics networks is reduced” (Chaberek 2002).

Regarding the logistics support only the savings formula is used because the basic process requires only a certain number and type of resources. In this situation, the possessed manufacturing potentiality of logistics services should demonstrate the greatest performance, since such action results in a decrease of single fixed costs, and the whole of actions becomes more effective (Chaberek 2002).

According to the literature ratings the macro-logistics system “(...) exhibits the highest level of aggregation and covers with its operation the entire national economy. It is characterized by the general economic term of flow of goods and information” (Nowicka-Skowron 2000). Macro-logistics, going further beyond the level of managing entities dealing with logistics management activities within the

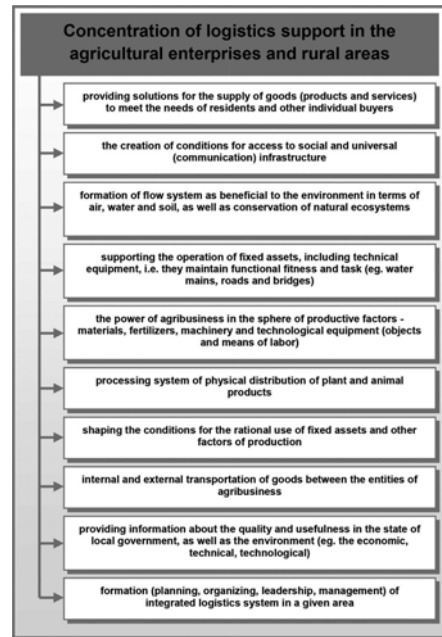


Figure 6. Concentration of logistics support in the agricultural enterprises and rural areas.

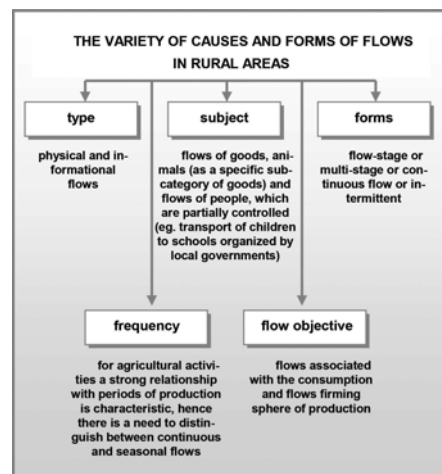


Figure 7. The variety of causes and forms of flows in rural areas.

given structures and legal administrative divisions, is treated within the economy as a whole. It is also emphasized that the macro-logistics system goes further outside the domestic economy, suggesting consideration in a broader context: “(...) in the case of logistics systems in terms of macro, they have a general economic nature. They may relate to, for example, logistics systems in the country or even around the world (global logistics)” (Grzybowska 2009).

Taking into account the above assumptions and statements and recognizing as correct quoted by M. Chaberek (2002) J. Dlugosz's suggestion regarding the identification of the fundamental surfaces revealing relationships within the logistics system, as well as between the logistics system and other systems of the changing environment, one may specify the basic elements of logistics support for agribusiness, grouped on three levels:

- internal—relationships within the logistics support system including physical flows and physical availability of resources;
- support for the supply and distribution of goods necessary in agricultural production and fulfilling the needs of rural areas;
- return proceeds (recycling and disposal) of waste resulting from the processes of production and consumption in rural areas;
- basic processes of the logistics chain: transportation, warehousing, storage, packaging;
- external—the relationship between the basic system and the system of logistics support carrying out the tasks arising from the demand for processes (and their products) occurring in the basic system: planning and designing the logistics support for single basic processes (eg. biomass production, bio fuel production, food production, agritourism development, etc.). Projecting and managing of networks of logistics operating the core processes in terms of their coverage (local—global); testing, measuring, handling operations related to the execution of transfers of cargo in the basic system; efficiency ensuring, reliability of equipment of basic system, as well as logistics support system (spare parts management, reorganization and modernization); exterior—the relationship between logistics system and its supporting systems to maintain the operation and efficiency: ensuring the mobility of human resources in the logistics support system; training support, training and improving of skills; supporting services related to eg. the maintenance of the logistics network (searching for cooperating entities, coordination, development of partnership within the support system, etc.). IT support for efficient implementation of the system of logistics support; technical documentation, information systems, databases used for managing and developing of the logistics support system.

4 CONCLUSION

The article has discussed the problems and barriers related to logistics management in agribusiness in terms of systemic approach, which referred to the raise of demand for implementation of activities

occurring in agribusiness and identification of challenges with regard to logistics support development. On the basis of implementation, one may conclude that every business organization should understandably be equipped with the system of logistics support. Therefore, the logistics support system reflects the fundamental processes of the organization and its structure, i.e. the elements and relationships between these elements.

Multitasking problem of economic cooperation due to the processes of globalization strictly requires logistics management functions supporting this cooperation. Ultimately, the integrated business management environment will be formed (Chaberek 2001). The management therefore takes place under conditions of severe relationship with the environment including all aspects of the economic, political as well as social world. Until recently, the entities have regarded their surroundings as static, and the events occurring in it as easy to predict. However, over the last twenty years of the twentieth century, not only in the world of politics, but also in social and economic world, the processes had been observed which completely changed the way of looking at business conditions. This tendency included also the theme of agribusiness, as a field of implementation of the principles and logistics management solutions aimed at shaping the system of flow of goods and information.

Raising demand for implementation of activities of systemic logistics management in the present world has a strong influence on agribusiness and states the fact that in current conditions, not only the municipality, but also the farmers are obliged to adapt to the changing circumstances of functioning in a competitive environment (Doh 2000). In order to ensure the development of the company (Malara 2008) and its resilient working, the entrepreneur needs to reach for the management, logistics and marketing solutions. The process of implementing demanded activities in the field of agribusiness means that farms are now regarded as enterprises and therefore optimal solutions in terms of systemic logistics management should be searched for.

REFERENCES

- Altekar R.V. 2005. *Supply Chain Management: Concepts and Cases*. PHI Learning Pvt. Ltd.
- Blaik P. 2001. *Logistyka*. PWE, Warszawa, p. 62, p. 261.
- Bozarth C. & Handfield R.B. 2007. *Wprowadzenie do zarządzania operacjami i łańcuchem dostaw*. Wyd. HELION, Gliwice.
- Bruska A. 2012. *Wsparcie logistyczne na obszarach wiejskich—istota i wyzwania*. Journal of Agribusiness and Rural Development, 3(25).

- Brzóska J. 2007. *Modele strategiczne przedsiębiorstw energetycznych*. Wydawnictwo Politechniki Śląskiej, Gliwice, p. 15.
- Chaberek G. 2002. *Funkcja wsparcia logistycznego w procesie globalizacji*, [in:] D. Perlo, P. Piątkowski (ed.), *Zarządzanie w warunkach globalizacji*. FPRP, Białystok, p. 47.
- Chaberek M. & Karwacka G. 2009. *Logistyka jako praktyczne urzeczywistnienie prakseologicznych zasad dobrej roboty*. Avcta Universitatis Nicolai Copernici, Ekonomia XL, Nauki Humanistyczno-Społeczne, Z. 391, Toruń, p. 8.
- Chaberek M. 2001. *Funkcje logistyki w procesie globalizacji*, [in:] E. Gołębska (ed.), *Eurologistyka droga do sukcesu firmy*. III Ogólnopolskie Warsztaty Logistyczne. Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań, p. 159.
- Chaberek M. 2002. *Rozwój kanałów logistycznych jako źródło ekonomicznych korzyści globalnych procesów gospodarczych*, [in:] D. Rucińska (ed.), *Dostosowanie polskiego transportu do Unii Europejskiej*. Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk, p. 223.
- Ciesielski M. 2008. *Łańcuch dostaw a strategie konkurencyjne*, [in:] Z. Foltynowicz, G. Jasiczak, J. Szyszka, *Towaroznawstwo, opakowania, logistyka*. Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań, pp. 246–247.
- Daft R.L., Kendrick M. & Vershinina N. 2010. *Management*. Hampshire, p.7.
- Doh J.P. 2000. *Entrepreneurial Privatization Strategies: Order of Entry and Local Partner Collaboration as Sources of Competitive Advantage*, Academy of Management Review, nr 3, pp. 555–571.
- Foltynowicz Z., Jasiczak J. & Szyszka G. 2008. *Towaroznawstwo, opakowania, logistyka*. Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań, pp. 229–230.
- Ghillyer A.W. 2009. *Management. A real world approach*. McGraw-Hill Higher Education, Boston, Burr Ridge, Dubuque, New York, San Francisco, St. Louis, Bangkok, Bogota, Caracas, Kuala Lumpur, Lisbon, London, Madrid, Mexico City, Milan, Montreal, New Delhi, Santiago, Seoul, Singapore, Sydney, Taipei, Toronto, pp. 10–11.
- Goldratt M. 1984. *The Goal*. The North River Press Publishing Corporation.
- Grzybowska K. 2009. *Podstawy logistyki*. Wyd. Difin, Warszawa, p. 48.
- Głód G. 2007. *Model tworzenia konkurencyjności wytwórców energii elektrycznej*, [in:] J. Pyka (ed.), *Szanse i zagrożenia rozwoju rynku energetycznego w Europie i Polsce*. Wydawnictwo Akademii Ekonomicznej w Katowicach, Katowice, pp. 96–97.
- Harrison A. & van Hoek R. 2010. *Zarządzanie logistyką*. PWE, Warszawa, p. 33–35.
- Hugos M. 2003. *Essentials of Supply Chain Management*. John Wiley & Sons, Hoboken, New Jersey, p. 2, p. 6, p. 17.
- Kearney A.T. 1978. *Measuring productivity in physical distribution*. National Council of Physical Distribution Management, p. 23.
- Ketikidis P.H. & Koh S.C.L. 2008. Dimitriadis N., Gunasekaran A. & Kehajova M., *The use of information systems for logistics and supply chain management In South East Europe: Current status and future direction*. Omega, p. 36.
- Kotler P. 2005. *Marketing*. Wydawnictwo Rebis, Poznań, p. 76.
- Larson P.D. & Halldorson A. 2004. *Logistics Versus Supply Chain Management: An International Survey*, International Journal of Logistics. Research and Applications, Vol. 7, p. 19.
- Malara Z. 2008. *Przedsiębiorstwo wobec wyzwań współczesności*, [in:] Z. Dworzecki, M. Romanowska (ed.), *Strategie przedsiębiorstw w otoczeniu globalnym*. Szkoła Główna Handlowa, Warszawa, pp. 31–32.
- Mangan J. Lalwani Ch. & Butcher T. 2008. *Global Logistics and Supply Chain Management*, John Wiley & Sons, Ltd; B. Mikus, *Strategisches Logistikmanagement. Ein markt-, prozess- und ressourcenorientiertes Konzept*, Deutscher Universitäts-Verlag/GWV Fachverlage GmbH, Wiesbaden 2003, p. 18.
- Markowski T. (red.). 2006. *Rola centrów logistycznych w rozwoju gospodarczym i przestrzennym kraju*. Biuletyn 255, KPZK PAN, Warszawa.
- Neider J. 2006. *Transport w handlu międzynarodowym*. Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk, s. 199–201.
- Nogalski B. 2011. *Modele biznesu jako narzędzie reorientacji strategicznej przedsiębiorstw*, [in:] W. Kieżun (pod red.), *Krytycznie i twórczo o zarządzaniu*. Oficyna a Wolters Kluwer business, Warszawa, p. 447, p. 451.
- Nowicka-Skowron M., *Efektywność systemów logistycznych*. PWE, Warszawa 2000, pp. 29–32.
- Oblój K. 2002. *Tworzywo skutecznych strategii. Na styku starych i nowych reguł konkurowania*, WN PWN, pp. 135–154.
- Pereira J.V. 2009. *The new supply chain's frontier: Information management*, International Journal of Information Management, 29, p. 373.
- Prockl G. 2007. *Logistik-Management im Spannungsfeld zwischen wissenschaftlicher Erklärung und praktischer Handlung*. Deutscher Universitäts-Verlag/GWV Fachverlage GmbH, Wiesbaden, pp. 369–425.
- Rutkowski K. 2004. *Zarządzanie łańcuchem dostaw - próba sprecyzowania terminu i określenia związków z logistyką*. Gospodarka Materiałowa i Logistyka nr 12.
- Shah J. 2009. *Supply Chain Management: Text and Cases*, Pearson Education India.
- Sople V.V. 2007. *Logistics management. The supply chain imperative*. Pearson Education. Delhi, p. 2.
- Sołtysik M. 2003. *Zarządzanie logistyczne*. Wydawnictwo Akademii Ekonomicznej im. Karola Adamickiego w Katowicach, Katowice, pp. 28–37.
- Stevens G.C. 1989. *Integrating the supply chain*, International Journal of Physical Distribution and Materials Management, Vol. 19, No. 9.
- Szymczak M. 2006. *Informatyzacja zarządzania logistycznego*, [in:] E. Gołębska (ed.), *Kompendium wiedzy o logistyce*. WN PWN, Warszawa, p. 158.
- Von der Gracht H.A. 2008. *The future of logistics scenarios for 2025*. Gabler Edition Wissenschaft, Wiesbaden, p. 88.
- Waters D. 2009. *Supply Chain Management*, Palgrave Macmillan, p. 9.

Implementation of the principles of the process orientation in the aspect of logistic management of supply chain

Dagmara Bubel

Częstochowa University of Technology, Częstochowa, Poland

ABSTRACT: Supply chain management involves integration of key business processes, from the end-user to suppliers of products, services and information that constitute added value for customers and other stakeholders. Within a company, activities are performed and coordinated, and every company is in some way engaged in relationships with other companies in a supply chain. The structure of activities within companies and between companies is of key importance for the productivity of the whole supply chain. Effective management of a supply chain requires integration of business processes with key participants of a supply chain. Valuable resources are wasted, when supply chains are not integrated, appropriately improved and managed. The value of standardised business processes results from the fact that managers from various organisations in a supply chain may use common language and link processes taking place in their companies to other members of the supply chain. The concept presented in this paper is based on key business processes, which run both in single enterprises and whole supply chains. Each process is managed by an interdisciplinary team which includes representatives of logistics, production, purchases, finances, marketing, research and development. Given that each process will cause interaction with key customers and suppliers, customer relationships management and supplier relationships management constitute critical links in the supply chain. The aim of this paper is to conduct a comparative analysis of various views on the type of processes that are present in the idea of supply chain management. Models designed for effective management of supply chain, such as Supply Chain Operations Reference (SCOR) and Global Supply Chain Forum (GSCF) will be analysed.

1 INTRODUCTION

In the 1990s, various authors attempted to express the nature of supply chain management in a single definition encompassing management philosophy, target group as well as objectives and numerous ways of their achievement. The subject of supply chain management is naturally the supply chain, which represents a network of organisations that are involved, through upstream and downstream supply chain links, in various processes and activities that create value in the form of products and services designed for end customer (Mesjasz-Lech 2013). In the broad sense, a supply chain consists of two or more legally separated organisations, connected with each other through a flow of materials, information and finances. Such organisations may include companies manufacturing parts, components and end products, logistic services providers, and even the end customer.

A network of connections usually does not focus on flows within a single chain, but it consists of comprehensive flows resulting from a number of different orders from customers that have to be processed parallel. As part of such comprehensive flows, a given organisation can concentrate only on one spe-

cific element of the supply chain. For instance, in the case of a downstream supply chain, organisations can be perceived as customers of their customers, whereas in the case of upstream supply chain, as suppliers of their suppliers (Müller 2015).

In the narrower sense, the term supply chain refers to large enterprises with a number of locations, often based in different countries. An effective coordination of the flows of materials, information and finances in such international companies still poses a huge challenge.

The aim of management of all the links in a supply chain is to increase competitiveness. This can be achieved, as individual organisational units are not individually responsible for competitiveness of their products and services in the eyes of the end customer, but the responsibility lies with the supply chain as a whole. Thus, competitiveness has become shifted from single companies towards supply chains. Naturally, in order to persuade a single company to become part of a supply chain it is presented with the prospect for a long-term win-win situation for all participants, which however does not have to apply to all entities in the short-term.

There are two possible ways to improve competitiveness of a supply chain. The first them is closer integration (or cooperation) of involved organisations, whereas the second one involves better coordination of the flow of materials, information and finances (Brzeziński, Brzozowska, Korombel 2014). Overcoming organisational barriers, adjustment of strategies and acceleration of the flows in a supply chain are the main issues in this context.

Given the above, supply chain management can be defined as a task of integration of the organisational units in a supply chain and coordination of the flows of materials, information and finances to meet (end) customer requirements for improvement of the competitiveness of the supply chain as a whole (Kauf 2015).

Studies conducted by researchers from Germany show (Heusler 2013; Hofmann 2014; Bogaschewsky et al. 2015) that management of supply chain involves integration of key business processes, from the end-user to suppliers of products, services and information that constitute added value for customers and other stakeholders.

2 CHARACTER AND SCOPE OF THE PROCESS ORIENTATION IN SUPPLY CHAIN MANAGEMENT

The idea of supply chain management is interpreted not only in the logistic context, but also in the integration and synchronisation views (Adamowicz, Lemanowicz 2013). However, regardless of how this concept is understood, its characteristic feature is process orientation, which means that decisions, actions and flows in a supply chain are treated as processes. In the process orientation, a supply chain is perceived as a sequence of internal and external processes which comprise a set of states and activities that serve a purpose of transition to the next states in which a delivered product has an increasing value. Within an enterprise, the problem is mainly to overcome functional barriers. This is usually the first step towards an extended enterprise where it is especially important to integrate processes going beyond the boundaries of the enterprises participating in the supply chain. This interpretation of supply chain management is expressed in the definition formulated by Werner and Wellbrock who claim that this is a strategic concept referring to understanding and managing a sequence of activities—from the supplier to the customer—that add value to supplied products (Werner 2015; Wellbrock 2015). However, based on the definition above, it is difficult to specify which processes should be integrated in accordance with the management concept discussed herein.

Since mid-1990s, there have been widely held views that cooperation in supply chain starts already at the stage of product creation and development. Other basic processes that should be managed at the level of a supply chain include demand planning and order fulfilment. Analysis of dependencies between a product and relationships occurring in supply chains shows an even larger scope of cooperation as part of supply chain management.

Analysis of relationships between product design and creation and network design and functioning enables distinction of four areas of supply chain management, namely (Sroka 2015):

1. Product and network configuration, which involves making key decisions concerning offered products and services, subject structure and connections between the links in the chain,
2. Product design using suppliers' potential of knowledge,
3. Formation of a production network aimed at selecting and defining production tasks, production and stock maintenance locations, which according to the deferring idea may refer not only to industrial enterprises but also commercial and logistic ones,
4. Optimisation of processes occurring in a supply chain which are connected with the physical flow of products and accompanying flows of information and financial resources.

Integration and coordination of decision-making processes connected with the above-mentioned areas of cooperation in a supply chain conform with idea of an “extended enterprise” (Adamczewski 2015), which has fluid boundaries and partners are perceived as components of an organisation. Second, they require cooperation between the participants of a supply chain which apart from logistic management includes joint planning as well as organisation and control of selected aspects of research and development, production and marketing activity. Clear assignment of decisions and activities to the individual areas of supply chain cooperation is not possible due to conventional and fluid boundaries between logistics, marketing and production management, which results in different scopes of their competences and responsibilities in organisational structures and rules of cooperating enterprises (Kadłubek 2015). One can only indicate that common decisions and activities at the borderline between logistics and marketing include: demand planning, design of distribution channels, care for the quality of customer service and cooperation to implement promotion campaigns. On the other hand, basic decisions and activities at the borderline between logistics and production management that are undertaken by the supply chain participants include: locating of

production facilities, assessment and selection of suppliers and stock definition and control.

3 MANAGEMENT OF LOGISTIC PROCESSES IN A SUPPLY CHAIN ACCORDING TO SCOR

Supply Chain Operations Reference (SCOR) was developed and popularised by members of the organisation Supply Chain Council that has been existing since 1996 (Poluha 2014). SCC not only defined the functional structure of software supporting supply chain management but it also provides principles and indicators for benchmarking of processes taking place within and between the links of a supply chain. The SCOR model enables measurement and adjustment of five types of processes in a supply chain: planning, purchase, creation, delivery and return of products [SCOR Model]. It is however not used for improvement of processes of product design and development, management of product sale and post-sale service, which essentially narrows its use to the sphere of logistics (Wannenwetsch 2014). In accordance with the assumptions of SCOR model, analysis and assessment of processes in a supply chain takes place at three hierarchically organised levels: process type, process category and process elements.

At the first of the above-mentioned levels, four types of real processes are distinguished (purchase, production, delivery and return) as well as a planning process for their coordination. At the second level of the model, they are decomposed into categories of planning, executive and protective and controlling processes. Moving to the third, most detailed level of the model, the individual categories of processes are divided into a logic sequence of actions and events comprising the processes. General ideogram of SCOR model has been presented in Fig. 1.

It should be stressed that the model's authors defined processes taking place in a supply chain in the most precise way and indicated a proposed set of indicators and metrics designed for its monitoring and assessment. Practical usefulness of SCOR

model lies in the fact that it can be used to describe, measure and assess supply chain processes. It is also a good tool that facilitates communication between the supply chain links, both at the stage of its designing and implementation and during the realisation.

4 PROCESSES IN SUPPLY CHAIN MANAGEMENT IN GSCD MODEL

Processes of supply chain management identified by The Global Supply Chain Forum basically refer to processes of managing relationships with customers, relationships with suppliers, customer service, demand, order fulfilment, production flow, product development and marketing and returns management (Wyrwich 2012). Each of the processes of supply chain management has both strategic and operational sub-processes. Strategic sub-processes ensure structure, i.e. how the process will be implemented, whereas operational sub-processes provide detailed steps for implementation. A strategic process is a necessary step in the integration of companies with other participants of a supply chain, but it is at the operational level that every day activities take place.

Customer relationships management defines the structure of how relations with customers will be created and maintained. A board of directors defines key customers and groups of target customers as part of a company's mission. The aim is to segment customers based on the value in time and to increase customer loyalty through provision of customizable products and services. Interdisciplinary teams prepare terms of Product and Service Agreements (PSA) to meet the needs of key customers and to segment the rest of customers. These terms specify levels of productivity. The teams work with key customers to improve processes and eliminate demand changeability as well as activities that do not create added value. Productivity reports are designed to measure profitability of individual customers as well as financial results for customers (Herold 2012).

Supplier relationships management is a process that defines how a company cooperates with suppliers. As the name itself suggests, it is a mirror reflection of customer relationships management. Just as a company needs to develop relations with customers, it also has to build ties with suppliers. Like in the case of customer relationships management, companies will create close relations with a small group of their suppliers, maintaining a remote contact with the rest. Product and service agreements are negotiated with every key supplier, defining the terms of the relationship. For segments of less important customers, terms of cooperation are not subject to negotiation. Supplier relationships management mainly refers to

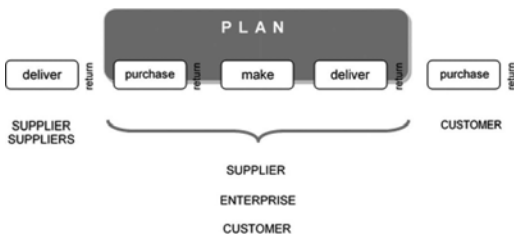


Figure 1. Five basic processes in SCOR model. Source: on work based on: Poluha 2014.

definition and management of these terms. Long-term relations are developed with a small group of main suppliers. The desirable result is the “win-win” situation where both parties achieve benefits.

One can say that customer service management is a company’s showcase for its customers. It is the main contact point for administration of product and service agreements. It provides customers with real time information about the arranged shipment date and product availability through interfaces with such company units as production or logistics. The process of customer service can also help the customer in the application of the product.

Demand management is a process in supply chain management that matches customer requirements with possibilities of a supply chain. Taking into account appropriate processes a board of directors can dynamically adjust the supply to the demand and develop a complex plan taking into consideration possible interruptions. This process is not limited to forecasting. It includes synchronisation of the supply and demand, increase in flexibility and decrease in demand changeability. A well-organised process of demand management can enable a company increased activity in the case of predicted demand and more effective reactions to unexpected changes in demand.

The process of order fulfilment is something more than order delivery. It includes all activities that are necessary to define customer requirements and design processes that allow a company to meet customer requirements minimising at the same time total costs of delivery and order fulfilment. This is not only a logistic function, it requires implementation of an interdisciplinary approach that goes beyond the boundaries of the individual departments and enables coordination of relations with key suppliers and customers. The aim is to create an efficient process from a supplier to an organisation providing services or products and to various segments of customers.

Management of production flow is a process in supply chain management that encompasses all activities necessary for the transfer of products through production processes in factories aimed at achievement and management of production flexibility in a supply chain. Production flexibility reflects ability to create a wide range of products in the right time, at possibly lowest costs. In order for a desirable level of flexibility to be achieved, the stages of production, planning and implementation have to go beyond the producer and occur at the level of a supply chain.

Product development and marketing is a process in supply chain management that provides organisational framework for developing products and bringing them to the market together with customers and suppliers. The tasks of a team responsible

for the process of product development and marketing include, in coordination with the process of customer relationships management, identification of articulated and not articulated customer needs. It is followed by selection of materials and suppliers in compliance with the process of supplier relationships management as well as development of production technologies and integration with the flow of goods in a supply chain—integration with the market.

Returns management is a process in supply chain management in which the object of management is activity connected with returns, reverse logistics, policy on returns within the organisation and towards key members of a supply chain. Proper implementation of this process enables not only effective management of reverse flow of goods but also identification of the possibilities to reduce the number of unwanted returns and control of reusable means, i.e. containers. Effective returns management is an important part of supply chain management and enables achievement of permanent competitive advantage.

5 PARTNERSHIP IN SUPPLY CHAIN AS INDICATOR OF PROCESS ORIENTATION

Partnership is adoption of trade relationships based on mutual trust, openness, risk and benefits sharing, which leads to companies achieving higher effectiveness than they would have achieved if they had cooperated not on the partnership basis (Frączkiewicz-Wronka, Szołtysek 2013).

An important aspect in the implementation of supply chain management is creation of appropriate relationships between participants of the chain. Both practitioners and the academic circles defend the importance of partnership, but it is still a challenge to find effective methods for the development of the proper type of relationships. In an environment characterised by limited resources, increased competition, higher customer expectations and faster pace of changes, managerial staff turn towards partnership in order to strengthen the integration of the supply chain and achieve permanent competitive advantage. Partnership is a way to use unique skills and experience of each of the partners and to eliminate competitors (Wellbrock 2015).

However, partnerships are expensive in terms of time and they take effort. Thus, an enterprise cannot and should not establish partnership with every supplier, customer or third-party service provider. It is important to make sure that the limited resources will be earmarked only for those relationships that will really bring benefits from the

partnership. However, many organisations get involved in relationships that do not meet the expectations of the board of directors and/or result in failure. It is thus important how managers can determine in advance whether the potential cooperation will lead to competitive advantage and is worth the time and resources necessary for full development into partnership. Moreover, not all partnerships are the same. It is important to know what type of partnership will ensure the best profits.

The model of partnership in supply chain proposed by Fröhlich, Buchta and Malilo (Fröhlich, Buchta, Malilo 2015) consists of four parts, i.e. factors influencing decisions to establish partnership, partnership mediators, partnership components and partnership effects and results, which have been presented in Figure 2.

Factors influencing decisions to establish partnership are convincing reasons for two companies related with expected profits of the established relationship. They can be encapsulated in four categories: asset/cost effectiveness, customer service improvement, marketing benefits and profit increase/stability. The stronger the factors, the bigger the chances for successful cooperation.

Mediators constitute an environmental factor which increases probability of partnership success. These factors exist in all types of business relations and cannot change in the short-run. Mediators determine what the potential of a network of partners is and they encompass (Matwiejczuk 2014):

- Compatibility of corporate cultures,
- Compatibility of management philosophies and techniques,
- Prospects for mutual cooperation between potential partners,
- Level of symmetry between enterprises.

Moreover, there are five factors regarded as mediators which can, but don't have to, influence the strengthening of partnership, namely (Schmäh, Gutsche, Meyer-Gossner 2015):

1. An enterprise has competitors' shares,
2. Enterprises are in close geographical proximity,

3. Possibility of exclusiveness as a result of partnership,
4. Previous experiences in relationships,
5. Enterprises have the same owner.

Components of partnership are managerial, controlled elements of partnership. It is thanks to the implementation of these components that the potential of partnership can be achieved. Components of partnership include (Streeck 2016):

- Style, level and content of planning,
- Metrics and controls of joint operations,
- Extent and type of communication between companies,
- Arrangements on how risk and benefits are distributed,
- Level of trust and involvement,
- Type of contracts used as part of partnership,
- Scope of activities between enterprises,
- Extent of joint investments.

Results reflect effectiveness of partnership and enterprises' ability to perform their statutory tasks. Results can be classified according to three basic categories, the first of which highlights global effectiveness results connected with improvement or rebalancing of profits, while the second determines such process results as improvement of service provision or cost reduction, and the third one determines competitive advantage referring to market positioning, market share or market knowledge.

6 ANALYSIS OF SUPPLY CHAIN OPERATIONS REFERENCE MODELS IN THE CONTEXT OF THEIR EFFECTIVENESS

The analysis of the nature, scope and structure of processes in supply chain management conducted herein confirms the ambiguity of this concept of management in the theoretical and methodological aspects. This is characteristic of many other modern concepts in management studies that are still being developed (Czakon 2015). However, theoretical ambiguity, different level of detailedness of models and ongoing discussion on the type and structure of processes comprising the idea of supply chain management do not have to limit its application values. Each of the models presented was developed based on empirical studies, and its effectiveness has been sufficiently verified in economic practice. Moreover, in each of the presented views, the condition of supply chain integration are assumptions about partnership, trust, informational transparency and appropriate distribution of risk and benefits among its links. Therefore,

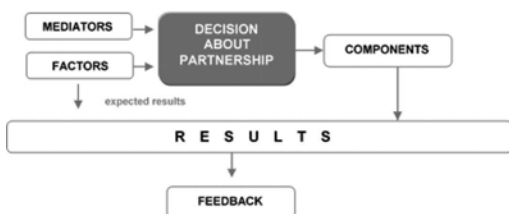


Figure 2. Model of partnership in a supply chain. Source: own work based on Fröhlich, Buchta, Malilo 2015.

managers, based on their knowledge and conviction, can choose any of the models designed for coordination and integration of processes taking place in a supply chain. So far, there have been no reliable research findings showing that their implementation leads to effectiveness or efficiency in a significant number of cases. Large scale research on the effectiveness of the SCOR model compared to other models of supply chain management is being carried out. For the time being, one can only say that findings of such research will allow the reference models presented herein to be better and widely used to improve processes taking place in supply chains and networks.

The conducted empirical studies allow to conclude that the structure and activities within and between companies form the foundation for creating and increasing effectiveness of supply chain. According to available studies (Richter-Von Hagen, Stucky 2013), managerial staff agree with the opinion that competitiveness and effectiveness may increase if key and internal business activities and processes are interlinked and managed along many enterprises. Thus, success of companies requires the introduction of changes from management of individual functions to integration of actions in processes of supply chains. Individual authors (Heusler 2013; Morita, Machuca, Flynn et al. 2015) propose implementation of business processes in the context of supply chain management, but so far there is no standard specifying what processes should be taken into account. The value resulting from choosing common business processes is that managers from organisations in a supply chain can use the same language and can link appropriate processes of their enterprises with other members of the chain.

It should be also stressed that most metrics called “supply chain indicators” (Sürrie, Reuter 2015) allow only to present dependencies occurring in the internal environment of an enterprise, not showing how profits or profitability in a supply chain are achieved. Going further, one can say that these metrics can introduce dysfunctions in a supply chain, as they are used to optimize results achieved by a given enterprise, but often at the cost of other enterprises from the supply chain. The use of the metrics discussed herein can thus lead to reduced effectiveness of the whole supply chain.

It is commonly believed that a well-developed system of metrics of a supply chain may increase chances for success by simultaneous adjustment of processes in a number of companies or by directing activity to most profitable market segments. It may also be a source of competitive advantage achieved through diversification of products/services and reduction of costs. On the other hand, an inappropriately prepared system of metrics will

cause difficulties in the fulfilment of consumer expectations and in competing in a supply chain as well as occurrence of apparent optimisation of a department's or company's productivity (Kauf 2015).

In order to avoid these threats, it is necessary to implement in the works on the development of supply chain metrics such frameworks that take into account increase in shareholders' profits and improvement of the process of customer relationships and supplier relationship management in each link of a supply chain. Combining process optimisation with supplier and customer profitability is the basis for creation of a system of metrics designed to identify chances for profitability improvement and definition of objectives by all companies from a specific supply chain. By identifying profitability sources in each of the links, managers of companies can make decisions that will enable productivity maximisation of the whole supply chain.

7 CONCLUSION

Management of processes is a necessary mechanism for optimal and efficient performance of processes within a given enterprise and the whole supply chain. The paper has reviewed the definitions of supply chain management provided by literature. It has outlined the concept of key processes of supply chain management proposed by researchers from the United States. It has highlighted both the strategic and operational character of each of the processes. Processes taking place at the level of a supply chain are closely connected with partnership between the individual participants of a supply chain. Therefore, the paper has paid attention to this issue taking into account its components. The assumptions of the concept of the partnership model in a supply chain have been outlined. Specificity of effectiveness measurement in a supply chain and necessity to select appropriate metrics have been indicated.

REFERENCES

- Adamczewski, P. 2015. Informatyczne wspomaganie organizacji sieciowych. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu* 402: 11–19.
- Adamowicz, M. & Lemanowicz, M. 2013. Koncepcja zarządzania łańcuchami dostaw versus tradycyjne pojmowanie relacji dostawca-odbiorca. *Economic and Regional Studies* 6 (1): 5–18.
- Bogaschewsky, R., Eßig, M., Lasch, R., & Stölzle W. (eds.). 2016. *Supply Management Research. Aktuelle Forschungsergebnisse 2015*. Wiesbaden: Springer Fachmedien.

- Brzeziński, S., Brzozowska, A., Korombel, A. 2014. Tools for integrating enterprises in a supply chain. Part 2. *Logistyka* 5: 28–30.
- Czakon, W. 2015. Sieci międzyorganizacyjne w naukach o zarządzaniu – w kierunku sieciowych modeli biznesu. *Studia Ekonomiczne* 217: 9–18.
- Frączkiewicz-Wronka, A., Szołtysek, J. 2013. Partnerstwo świadczące usługi społeczne w świetle doświadczeń logistyki społecznej. *Studia Ekonomiczne* 168: 281–292.
- Fröhlich E., Buchta Ch., Malilo N. 2015. Zur Integration von Nachhaltigkeitsrisiken in das Strategische Beschaffungsmanagement. In E. Fröhlich (ed.), *CSR und Beschaffung. Theoretische wie praktische Implikationen eines nachhaltigen Beschaffungsprozessmodells*: 55–75. Berlin Heidelberg: Springer.
- Herold, L. 2012. *Kundenorientierte Prozesssteuerung in der Automobilindustrie: Die Rolle von Logistik und Logistikcontrolling im Prozess "vom Kunden bis zum Kunden"*. Wiesbaden: Springer-Verlag.
- Heusler, K.F. 2013. *Implementierung von Supply Chain Management: Kompetenzorientierte Analyse aus der Perspektive eines Netzwerkakteurs*. Wiesbaden: Springer Fachmedien GmbH.
- Hofmann, E. 2014. *Interorganizational Operations Management*. Wiesbaden: Springer Fachmedien.
- Kauf, S. 2015. Wybrane atrybuty zarządzania publicznym łańcuchem dostaw – orientacja na klienta i przepływy. *Studia Ekonomiczne* 217: 56–67.
- Kauf, S. 2015. Zarządzanie łańcuchem dostaw w sektorze publicznym. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu* 383: 50–59.
- Kadhuk, M. 2015. The Selected Areas of E-logistics in Polish E-commerce. *Procedia Computer Science* 65: 1059–1065.
- Matwiejczuk, R. 2014. Z badań nad oddziaływaniem kompetencji logistyki na tworzenie przewagi konkurencyjnej przedsiębiorstwa. Pozycjonowanie i integracja. *Gospodarka Materialowa i Logistyka* 11: 2–9.
- Mesjasz-Lech, A. 2013. Wykorzystanie technologii informacyjnych w zarządzaniu łańcuchem dostaw. *Zeszyty Naukowe Uniwersytetu Szczecińskiego. Ekonomiczne Problemy Usług* 105: 543–552.
- Morita, M., Machuca, J.A., Flynn, E.J., & de los Ríos, J.L.P. 2015. Aligning product characteristics and the supply chain process – A normative perspective. *International Journal of Production Economics* 161: 228–241.
- Müller, M. 2015. *Informationstransfer im supply chain management: Analyse aus Sicht der Neuen Institutionenökonomie*. Wiesbaden: Springer-Verlag.
- Poluha, R. G. 2014. *Anwendung des SCOR-Modells zur Analyse der Supply Chain: explorative empirische Untersuchung von Unternehmen aus Europa, Nordamerika und Asien*. Lohmar—Köln: Josef Eul Verlag GmbH.
- Richter-Von Hagen, C., Stucky, W. 2013. *Business-Process- und Workflow-Management: Prozessverbesserung durch Prozess-Management*. Stuttgart—Leipzig—Wiesbaden: G.B. Teubner Verlag.
- Schmäh, M., Gutsche, J., & Meyer-Gossner, M. 2015. Wie Soft Skills und Social Business das Account Management revolutionieren. In S. Fließ, M. Haase, F. Jacob & M. Ehret (eds), *Kundenintegration und Leistungslehre: Integrative Wertschöpfung in Dienstleistungen, Solutions und Entrepreneurship*: 405–425. Wiesbaden: Springer Fachmedien.
- Sroka, W. 2015. Sieci logistyczne: wybrane aspekty tworzenia i funkcjonowania. *Studia Ekonomiczne* 217: 44–55.
- Streeck, W. 2016. Von Konflikt ohne Partnerschaft zu Partnerschaft ohne Konflikt: Industrielle Beziehungen in Deutschland. *Industrielle Beziehungen-Zeitschrift für Arbeit, Organisation und Management-The German Journal of Industrial Relations* 23(1): 47–60.
- Sürle, Ch., Reuter, B. 2015. Supply Chain Analysis. In H. Stadler, Ch. Kilger, H. Meyr (eds), *Supply Chain Management and Advanced Planning: Concepts, Models, Software, and Case Studies. Springer Texts in Business and Economics Supply Chain Management and Advanced Planning*: 29–54. Berlin Heidelberg: Springer.
- Wannenwetsch, H. 2014. *Integrierte Materialwirtschaft, Logistik und Beschaffung*. Berlin Heidelberg: Springer.
- Wellbrock, W. 2015. *Innovative Supply-Chain-Management-Konzepte: Branchenübergreifende Bedarfsanalyse sowie Konzipierung eines Entwicklungsprozessmodells*. Wiesbaden: Springer Fachmedien.
- Werner, H. 2015. Supply Chain Performance messen. *Controlling & Management Review* 59 (1): 18–25.
- Wyrwich, S. 2012. Zarządzanie łańcuchem dostaw – wyzwania w zakresie nowoczesnych form pracy. *Logistyka* 5: 246–251.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Analytical grounds and effective operation area of green logistics management in the transport area

M. Kadłubek & K. Grondys

Czestochowa University of Technology, Czestochowa, Poland

ABSTRACT: The scope of green logistics matters is increasing as the theory and practice of its functions, processes and relationships are on the rise. The most important area of green logistics is transport as the one with the huge impact on environment. Elimination of contamination, transfer of high volume freight moving from roads to rail, preservation of appropriate environmental management standards are valuable factors in term of green logistics in transport area. Analysis of the nowadays importance of green logistics and its transport area is the aim of the article. Particular attentiveness was directed to the proposals of building proper analytical grounds as well as green logistics effective operation area.

1 INTRODUCTION

In reference to the World Economic Forum and Accenture (2009) data, logistics determines about 5.5% of greenhouse gas emissions of the world, which is designated to the division between modes of freight transport and logistics buildings. Greenhouse gas emissions of the world with the origin in logistics with its footprint seems to be fairly unpretentious, however, in contrast to a large amount of the other sectors which need to reduce greenhouse emissions, the transport sector, specially freight transport increases its productivity of these gases.

In many industries the term green logistics is used to refer to implementation of proactive environmental protection strategy (Romanowska 2010) and accomplishment of operational decisions. In the area of transportation the companies are under the pressure to develop costly responsible and effective activities which will also be in correlation with environmental awareness and green logistics solutions. The aim of the paper is analysis of the nowadays importance of green logistics and its transport area management with particular emphasis on possibilities of building proper analytical ground and green logistics effective operation area.

2 FIELD OF GREEN LOGISTICS AND ITS TRANSPORT AREA

Nature of logistics is functionally related, intensely integrative and various logistics activities influence the environment. Environmental objectives are closely aligned to economic objectives in the sphere of logistics designated as green logistics.

Mesjasz-Lech (2012) points out that green logistics includes the whole of activities interrelated to the eco-efficient management of the flows of products and information from the point of origin to the point of consumption with the aim to meet or exceed customers demand at minimum cost. In other words green logistics term combines the environmental attributes and logistics activities to manage them in the approach which take into consideration the importance of environment in all decision making activities, operations and processes of logistics (Pishvee et al. 2012). On the foundations of acknowledgment that respective enterprises environmental influence expands well further than their companies borders, Klassen & Johnson (2004) settled their vision of green supply chain (Brzeziński, Brzozowska, Korombel 2014) management as the arrangement and integration of environmental management within supply chain management. Aksoy (et al. 2014) identify green logistics with producing and distributing products in sustainable way, taking into account the environmental and social factors. Green logistics forces all users logistics system, to consider how their actions affect the environment. The main objective of green logistics is to coordinate all activities in such a way that while the use of their supply chains in the most efficient way, minimizing the cost was going to be borne by the environment (Bajdor 2012).

The extent of green logistics themes is growing as the theory and practice of its functions, processes and relationships are developing. The most widespread and frequent themes of green logistics due to McKinnon (et al. 2015) are:

- Reducing freight transport externalities,
- City logistics,

- Reverse logistics,
- Logistics in corporate environmental strategies,
- Green supply chain management.

From economical and managerial perception, the logistics and specially transportation area commonly are in contradiction with sustainable propose of logistics and environmental responsibility related to green logistics (Kadłubek 2015).

Nowadays since sustainability is becoming more and more imperative business feature (Grudzewski et al. 2010), the consequences are also in building proper analytical grounds looking for the measures and methods to facilitate evaluation of the complete image of environmental effects related to the activities, also in green logistics and transport area.

3 PROPOSALS OF ANALYTICAL GROUNDS FOR GREEN LOGISTICS MANAGEMENT IN THE TRANSPORT AREA

Green logistics activities include measuring the environmental impact of logistics areas. As the freight transport on average determines about 80–90% of carbon dioxide emissions related to logistics, this is the main reason why it is also the major area of carbon-reduction labors and other activities focused on identification and improvement its environmental impact. These activities should take into account environmental issues inte-

grated with logistics in the transport area in order to change the environmental performance, reduce negative effects and create eco-friendly solutions. First of all they have to be properly determined and analyzed, therefore measurements of this area appear to be initial in constructing the methodical basement. Few proposals of formations of the analytical grounds for green logistics in the transport area are introduced below.

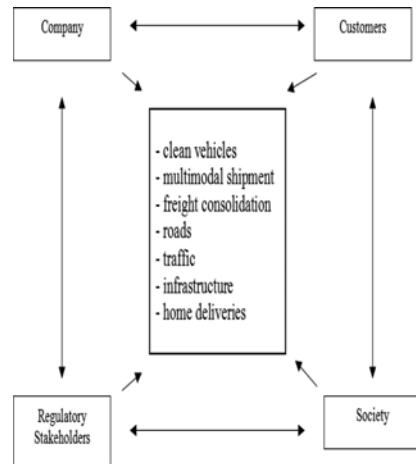


Figure 1. Four factors affecting green logistics management and indications for the transport area. Source: Own elaboration on the basement of: (Schmied 2010).

Table 1. Parameters for analysis related to the green logistics management in the transport area.

Parameter	Depiction
Modal split	Indicates the proportion of freight carried by different transport modes and can be expressed as the ratio of tonne-kms carried by more carbon-intensive modes such as road and air to tonne-kms carried by greener modes like rail, barge, ship and pipeline
Average handling factor	Ratio of the weight of goods in an economy to freight tones-lifted, as products passed through the supply chain are loaded often several times
Average length of haul	Mean length of each link in the supply chain; coverts the tonnes-lifted statistic into tonne-kms
Vehicle utilization	Can be measured by the ratio of vehicle-kms to tonne-kms, in other words how much vehicle traffic is required to handle a given amount of freight movement; if the vehicles are well-loaded on outbound and return journeys this ratio is minimized
Energy efficiency	The ratio of energy consumed to vehicle-kms travelled; it is a function mainly of vehicle characteristics, driving behaviour and traffic conditions
Emissions per unit of energy	The amount of CO ₂ and noxious gases emitted per unit of energy consumed either directly by the vehicle or indirectly at the primary energy source for electrically-powered freight transport operations
Other externalities per vehicle-km and per unit of throughput	Environmental effects such as noise irritation, vibration, accidents etc.
Monetary valuation of externalities	Physical measures of logistics-related externalities converted into monetary values

Source: Own elaboration on the basement of: (McKinnon et al. 2015).

Schmied (2010) makes a distinction of four general factors which affect green logistics and precise the most important indications for the analysis of the transport area. The factors are: company, customers, regulatory stakeholders and society, which in their environmental consciousness have requirements due to the transportation for home deliveries of the products with clean vehicles, safe for the green nature, developing operations of multimodal shipment and freight consolidation with the aim of emissions reduction, and other eco-friendly solutions for transport infrastructure, roads or traffic (Figure 1). These indications in the transport area should be essential impose for measure assortment in green logistics.

McKinnon (et al. 2015) propose key parameters in analysis related to green logistics management in the transport area, which are presented in Table 1.

In the context of sustainable development, the transport area involves three dimensions while building analytical ground: economic, social and environmental dimensions. In reference to Russo & Comi (2012) typically used measures are focused on several vital aspects with examples indicated in Table 2.

In the category of logistics systems, universal cost trade-offs involving areas of transport, warehousing and inventory are also sensitive to environmental impact and greenhouse gas. Orienting these cost trade-offs adequately to stimulate a direction to more decentralised outlines of production area and supply chain, would force huge extension in freight transport costs. Establishing of the carbon dioxide influence is also possible by accomplishing

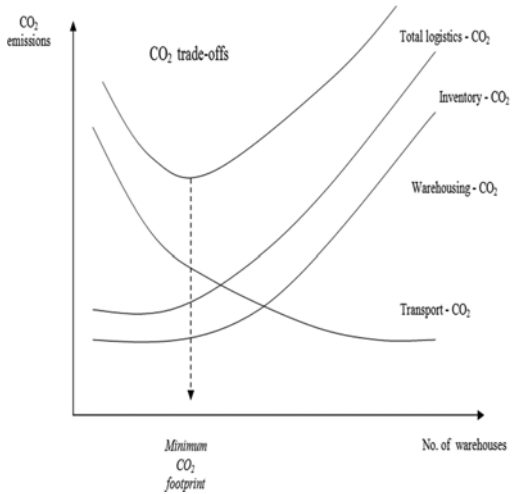


Figure 2. Logistic system with respect to CO₂ emissions as relation to green logistics. Source: (McKinnon 2010).

a green logistics trade-off analysis, along the theoretical and practical background which is functional in the management optimization of logistics systems, but related to the respect to CO₂ emissions (Figure 2).

4 GREEN LOGISTICS EFFECTIVE OPERATION AREA FOR TRANSPORTATION

Over the past years green logistics has represented a lot of nature trails, and one of the most distinguishable is reduction of transport costs by choosing cheaper and more eco-effective type of transport, e.g. rail transport. Unfortunately in recent years, significantly increased the share of road transport in the provision of green logistics services (Bubel 2015). The fact that the road transport has a number of advantages, such as the ability to deliver small consignments “door to door”, which is a important transport logistics technology, is irrevocable. On the other hand it should also be noted that each mode of transport has its own optimum range of action. For short distances (150–200 km), small cargo transportation may be difficult and often even impossible for the railway transport to compete with road transport. But in many cases the choice between the mode of transport is less obvious. Then the premises of the companies to realize green logistics management activities may be primarily motivated by environmental considerations while taking into the consideration cost effectiveness of the operations

Table 2. Dimensions and their aspects for building analytical ground of transport area in the context of sustainable development.

Dimension	Aspect
Economic	<ul style="list-style-type: none"> – route length – delivery times – infrastructural costs – transport congestion (i.a. additional times spent in journey, journey time, journey speed)
Social	<ul style="list-style-type: none"> – decreasing interference between individual segments of mobility – decreasing number of vehicles engaged in their tasks – decreasing number of traffic accidents
Environmental	<ul style="list-style-type: none"> – decreasing pollution – decreasing the noise level – loss of residential space

Source: Own elaboration on the basement of: (Russo & Comi 2012).

(Man & Nowicka-Skowron 2010). Somehow in the range of e.g. rail or road transport effective operation is determined by the transport service change point (CP), which is calculated by the formula (Brusyanin et al. 2013):

$$CP = \frac{F_1 - F_2}{V_2 - V_1}, \quad (1)$$

where:

- F_1 and F_2 are fixed costs of the first and the second type of transport;
- V_1 and V_2 are variable costs of the transport types.

The graph below (Figure 3) shows that the area within the range of 0 to the point CP may be more or less cost-effective in accordance to first or second type of transport. The problem of green logistics to reduce the adverse impact of transport on the environment may be transformed into a problem of various types of transport integration, the implementation of their cooperation with a minimum of vehicles, particularly into the task of the organization of intermodal transport.

where:

$F_{1,2}$ —the fixed costs of the first and second types of transport;

$V_{1,2}$ —the variable costs of the first and second types of transport;

$F+V$ —the overall cost of transport.

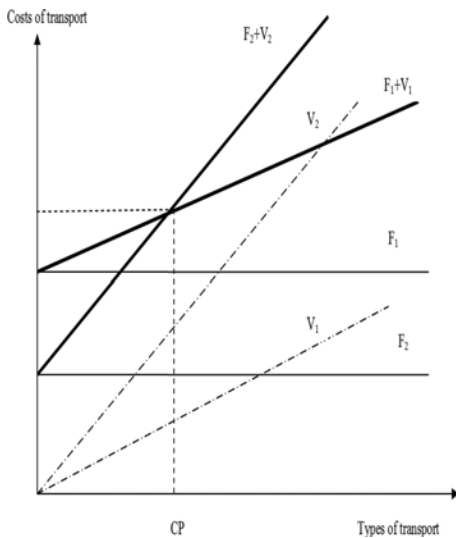


Figure 3. Correlation between the transport costs when choosing a type of transport. Source: (Brusyanin et al. 2013).

In reference to the intend to determine the area of effective operation of the transport types, the proposal of the area called “the GreenLogistics” may be introduced as below. To realize it, into the formula (1) one needs to insert the green area, followed by the development of a fine system for each type of transport, both on variable and fixed costs or the green start (green benefit). Below it is proposed to designate it with the same term: “Green Logistics”. If the “GreenLogistics” (fine) has a value greater than 1 and “GreenLogistics” (benefit) is located within the range of 0 to 1 as below:

GreenLogistics > 1 ⇒ Area

$$0 < \text{GreenLogistics} \leq 1 \Rightarrow \text{Start}, \quad (2)$$

then the amount of fines-benefits of each type of transport will give a coefficient (called “green”); the calculation results are presented in Table 3.

Then the formula (1) takes the form:

$$CP = \frac{a_1 F_1 - a_2 F_2}{b_2 V_2 - b_1 V_1}, \quad (3)$$

where:

a_1, a_2 —“GreenLogistics” coefficient of fixed costs of the first and second transport type:

$$a_{1,2} = \sum_{j=1}^k \text{GreenLogistics}_j, \quad (4)$$

where:

b_1, b_2 —“GreenLogistics” coefficient of variable costs of the first and second type of transport:

$$b_{1,2} = \sum_{j=1}^k \text{GreenLogistics}_j, \quad (5)$$

Getting parameter values in the right-hand sides of the formulas (4) and (5) can be the subject of a separate study. They can be determined

Table 3. The calculation of “GreenLogistics” coefficients for determining effective operation area of various types of transport.

GreenLogistics		Fixed costs, F_i	Variable costs, V_i
Benefit	GreenLogistics ₁
	GreenLogistics ₋₁
Fine	GreenLogistics _j
	GreenLogistics _k
Total		a_i	b_i

by the expert assessments method or mathematical modeling (Kazakov et al. 2011). One can visualize how the transport service change point (CP) can be changed while implementing the “GreenLogistics” by using the example graph as below (Figure 4).

where:

F_{2*} —the fixed costs of the second type of transport with the implementation of green fine;

CP^* —the transport service change point with the implementation of green fine for fixed costs.

where:

V_{2*} —the variable costs of the second type of transport with the implementation of green fine;

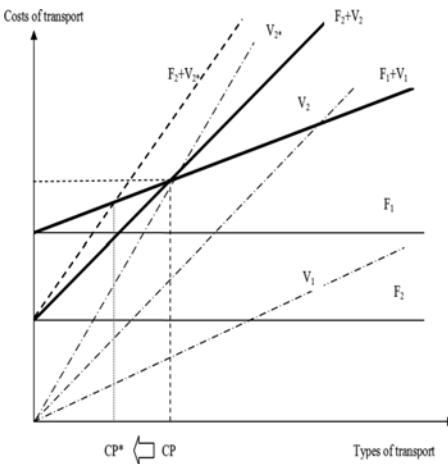
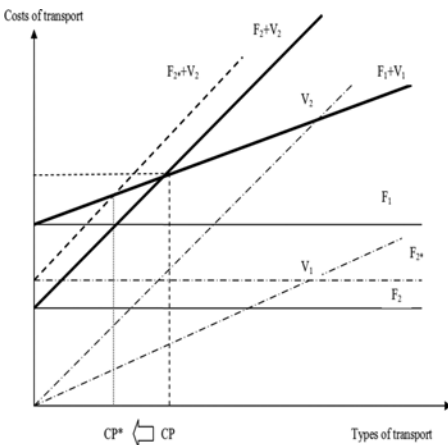


Figure 4. Examples of shifting the transport service Change Point (CP) with the implementation of green fines for fixed costs (a) and variable costs (b) of the transport types. Source: Own elaboration on the basement of: (Kazakov et al. 2013).

CP^* —the transport service change point with the implementation of green fine for variable costs.

The above example graphs show that with the implementation of the “GreenLogistics” effective operation area of the variable costs of the second type of transport is sharply reduced, and the effective operation area of the fixed costs of the second type of transport is increased. This approach to the transport types evaluation in terms of their effectiveness seem to be efficient in the use of the principles of green logistics management as an obligatory direction of development. Accordingly, proposed path in the beginning of creation of “GreenLogistics” area, based on the promotion of transport types implementing green technologies and fining those types of transport that do not pay enough attention to environmental issues, may allow to estimate the contribution of each type of transport in the solution of environmental problems for the benefit of future generations.

5 CONCLUSIONS

In view of profitability aspects as a primary requirement in order to exist in nowadays exceedingly growing competition, green logistics, specially its transport area and its management, has significant position in the national and international sustainability objectives (Nogalski & Szpitter 2014) since they essentially are inclined towards environmentally friendly transportation types and modes. Elimination of contamination, transfer of high volume freight moving from roads to rail, preservation of appropriate environmental management standards are valuable factors in term of green logistics management in transport area (Altuntas & Tuna 2013). Realization its aims, among others through building the proper analytical ground of the area, measures or searching for effective operations areas, moreover follow the line of international conventions concerning i.a. sustainable development (WCED 1987), Kyoto Protocol (UNFCCC 2012), European Union initiatives as White Paper on transport (European Commission 2011) and others.

Carbon dioxide emissions determined by logistics should be reduced due to numerous negative consequences, specially in the sphere of transport. Reduction possibilities are also in the areas of logistics management operations with the beginning in their analysis with the use of right parameters of the carbon dioxide emissions related to the transport logistics. The proper decarbonisation analysis, measures or methods should as diminish emissions, as reduce the costs, produce flow of financial and ecological benefits.

Carbon intensity differs extensively among transport types and modes. The shift of freight from types and modes of transport with comparatively high carbon amounts, as air and road transport, to those with much poorer carbon dioxide emissions, as rail and water-borne transport, can considerably decarbonise freight transport operations. This is also the reason to follow the other possible solutions in realization of green logistics objectives. Such a proposal may be determination of the area of effective operation of the transport types, which may assist decision makers and managers acquire an in-depth understanding of environmental impacts and costs associated.

REFERENCES

- Aksoy, A., Kucukoglu, I., Ene, S. & Ozturk, N. 2014. Integrated emission and fuel consumption calculation model for green supply chain management. *Procedia—Social and Behavioral Sciences* 109: 1106–1109.
- Altuntas, C. & Tuna, O. 2013. Greening Logistics Centers: The Evolution of Industrial Buying Criteria Towards Green. *The Asian Journal of Shipping and Logistics* 29(1): 59–80.
- Bajdor, P. 2012. Comparison between sustainable development concept and green logistics—the literature review. *Polish Journal of Management Studies* 5: 236–244.
- Brusyanin, D.A., Say, V. M. & Vikharev, S.V. 2013. Validation of vehicles choice for the route of regular passenger road and rail transport. *Herald of USRT* 1(17): 50–64.
- Brzeziński, S., Brzozowska, A. & Korombel A. 2014. Tools for integrating enterprises in a supply chain Part 1,2. *Logistyka* 4,5.
- Bubel, D. 2015. Configuration of Flexibility of Logistic Services. *AD ALTA Journal of Interdisciplinary Research* 5: 10–16.
- European Commission. 2011. *White Paper. Roadmap to a Single European Transport Area—Towards a competitive and resource efficient transport system*. Brussels: European Commission.
- Grudzewski, W. M., Hejduk I.K., Sankowska A. & Wańtuchovicz M. 2010. *Sustainability w biznesie, czyli przedsiębiorstwo przyszłości. Zmiany paradygmatów i koncepcji zarządzania*. Warszawa: Poltext.
- Kadłubek, M. 2015. Transport Sector and the Assumptions of Low-Carbon Transformation of Poland. In B. Skowron-Grabowska (ed.), *Logistics and Marketing Determinants of Enterprises Management*. Ostrava: Vysoka Skola Banska—Technicka Univerzita Ostrava.
- Kazakov, A.L., Lempert, A.A. & Bukharov, D.S. 2013. On segmenting logistical zones for servicing continuously developed consumers. *Automation and Remote Control* 74(6): 968–977.
- Klassen, R.D. & Johnson, F. 2004. The green supply chain. In S.J. New & R. Westbrook (eds.), *Understanding Supply Chains: Concepts, critiques and futures*. Oxford: Oxford University Press.
- Man, M. & Nowicka-Skowron, M. 2010. Costs related to the functions of company logistics. *Polish Journal of Management Studies* 1: 23–33.
- McKinnon, A., Browne, M., Piecyk, M. & Whiteing, A. 2015. *Green Logistics. Improving the environmental sustainability of logistics*. UK: Kogan Page.
- McKinnon, A.C. 2010. Green Logistics: the Carbon Agenda. *LogForum* 6(3), 1: 1–9.
- Mesjasz-Lech, A. 2012. *Efektywność ekonomiczna i sprawność ekologiczna logistyki zwrótej*. Częstochowa: Wydawnictwo Politechniki Częstochowskiej.
- Nogalski, B. & Szpitter, A. 2014. Koncepcja sustainability jako determinanta rozwoju przedsiębiorstwa. In I. Hejduk & A. Herman (eds.), *Dla przyszłości*. Warszawa: Difin.
- Pishvae, M.S., Torabi, S.A. & Razmi, J. 2012. Credibility-based fuzzy mathematical programming model for green logistics design under uncertainty. *Computers & Industrial Engineering* 62: 624–632.
- Romanowska, M. 2010. Przełomy strategiczne w przedsiębiorstwie. *Studia i Prace Kolegium Zarządzania i Finansów SGH* 98: 7–15.
- Russo, F. & Comi, A. 2012. City characteristics and urban goods movements: A way to environmental transportation system in a sustainable city. *Procedia. Social and Behavioral Sciences*, 39: 61–73.
- Schmied, M. 2010. *Aktuelle Entwicklungen zur Standardisierung der CO₂-Berechnung*. Hannover: Institute for Applied Ecology.
- United Nations Framework Convention on Climate Change (UNFCCC). 2012. *Kyoto Protocol*. <http://unfccc.int>. Visited on 15.01.2016.
- World Commission on Environment and Development (WCED). 1987. *Our Common Future*. Oxford: Oxford University Press.
- World Economic Forum/Accenture. 2009. *Supply Chain Decarbonisation: the Role of Logistics and Transport in Reducing Supply Chain Carbon Emissions*. Geneva: World Economic Forum/Accenture.

Information and communication support for the agricultural sector of Ukraine

A. Kalinichenko

Opole University, Opole, Poland

O. Chekhlatyi

Poltava State Agrarian Academy, Poltava, Ukraine

ABSTRACT: In the article, the current state of information and communication provision of the agricultural sector of Ukraine is studied; the main results of the information and consulting services activity in Poltava region are represented; their contribution to the spread of the new information technologies in the agricultural sector of the country economy is demonstrated.

1 FORMULATION OF THE PROBLEM

Under market conditions, the issue of the development of communication technologies is highly important for the agricultural sector of Ukraine, as possessing information and its use in the production process organization is directly related to ensuring the food security of the state.

Rural area development is largely determined by applying more advanced forms of management that provide efficient use of economic mechanism under specific production conditions. For this purpose, diverse informational means of processing and analysis must be considered. Currently information, transmission, software and hardware processing are the equally important resources as material and energy. Information transmission in the system of management of enterprises in the agricultural sector requires effective organization (Khudyakov 2016).

Therefore, improving the quality of information and communication software can play a crucial role in improving the efficiency of agrarian enterprises of Ukraine. It will enable a more clear focus in the legislative area, forecasted rates of production and marketing, geography of prices for the products and the resources in order to define a strategy for economic development, implement and use new technologies, assure production, storage and sales and build financial relationships in a tactically proper way (Khudyakov 2016).

2 ANALYSIS OF RECENT RESEARCH AND PUBLICATIONS

Under current conditions, one of the crucial factors in the effective development of the agricultural sector is the design of an effective system of

information and communication security. Theoretical and methodological basis of the agricultural production informatization, formation of information security and its role in the management of the agricultural sector, foundation of information and consulting services were covered in the works of the ukrainian scholars O.A. Halych, M.F. Bezkrivnyi, V.V. Derlemenko, T.P. Kalna-Dubynyuk, I.M. Kryvoruchko, M.F. Kropyvko, and others. However, systematic study on the formation of the information and communication provision of the agricultural sector as well as the analysis of information needs of an entity has not been covered.

The importance of the abovementioned issues, the need for their thorough theoretical study and practical specification predetermined the topicality of the subject of our study, its goals and objectives.

The main purpose of the article is to study the current operating conditions of the information and communication provision in Ukraine, underscore the main results of the information and consultancy services in Poltava region, their international cooperation and contribution to the spread of new information technologies in the agricultural sector.

3 THE MAIN MATERIAL OF THE RESEARCH

The implementation of the Information and Communication Technologies (ICTs) and their widespread use in various areas of human life, society and state is one of the most important tools that promotes the increase of the level of economic, social, cultural and technological development (Larin & Rudenko 2013).

Taking into consideration the global trends, in the last ten years, Ukraine embarked on the development of the Information Society; it was confirmed in a number of conceptual and strategic documents, primarily in the Law of Ukraine “On the Fundamentals of Information Society in Ukraine for 2007–2015” (Larin & Rudenko 2013). At the same time, a set of unresolved issues of legal, organizational, technical, scientific and methodological, analytical, resource support for the information society development remains. A large number of government decisions on these issues has non-systematic and declarative character, is not financially supported and is largely “borrowed” from other countries without considering the peculiarities of the current state and trends of Ukraine. The official confirmation of this point of view is the annual report on the state of information and communication provision in Ukraine, which is developed and submitted for the Parliament by the government along with the draft of the state budget for the next year according to the National Informatization Program. It is necessary to underscore that this document has some shortcomings in its formulation and implementation, namely:

- lack of the study on the documents and the real impact on the budget process;
- incomplete and inaccurate data of the State Statistics Committee and other bodies of the public authorities included;
- delay in the consideration of the document in Parliament;
- ignoring the procedures of the strategic monitoring, analyzing and forecasting the state in the process of its formation.

Despite the drawbacks mentioned it performs a very important function, namely, informing citizens, society and state about the situation in this area, the main factors and trends that are influential, short-term priorities of the state policy (Larin & Rudenko 2013).

Due to the economic crisis and the war in Ukraine, in the last year there was a decrease in the rate of computerization of enterprises in various sectors of economy; financing of the information projects from the state budget has decreased almost in two times and the National Informatization Program in ten times (Larin & Rudenko 2013).

Therefore, according to the World Economic Forum, the ratings of Ukraine which were found in the ICT indexes are the following:

- WEF Global Competitiveness Index 2015—the 79th place out of 140 countries (in 2014, it was the 76th out of 144 countries);
- WEF Networked Readiness Index 2015 – the 71st place out of 143 countries (in 2014, it was the 81st place out of 148 countries);

- WEF Technological Readiness Index 2015 – the 86th out of 140 countries (in 2014, it was the 94th place out of 148 countries).

In correspondence to the UN Global E-Government Development Index, in 2014, Ukraine was ranked the 87th place in the world out of 193 UN member states (in 2012, in the ranking, Ukraine was on the 68th place out of 190 countries). Despite the loss of positions in the rankings, in particular, in the index of online services, Ukraine joined the group of countries with a high index of e-government in 2014 that is considered to be a positive trend for the country (Information Society 2016).

According to the report of the International Telecommunication Union “Measuring the Information Society 2015”, which includes the rankings of 167 countries, in the index of ICT development, Ukraine was rated the 79th place (according to the ITU Report for 2014, it was the 73rd one out of 166 countries).

One of the reasons for the low rate of Ukraine in that ranking is an uneven access to ICT in the regions that is confirmed by the results of the analysis on the development of information and communication infrastructure and ICT in different areas of the region.

According to the data of the International Organization World Wide Web Foundation on the Internet development ranking in 2014, Ukraine was given the 46th place out of 86 countries in the Web Index.

According to the annual report “The State of Broadband 2015”, prepared by a joint initiative of the International Telecommunication Union and UNESCO, the level of Internet penetration in Ukraine was ranked the 95th place out of 191 countries (in 2013, it was the 94th place out of 191 countries) (Information Society 2016).

Unfortunately, nowadays in Ukraine, both a strategy and effective mechanisms for the information society development are absent. In the process of the formation of these mechanisms, the global trends in ICT development, the peculiarities of Ukraine in the area of information and information society, modern approaches and methods of government (Larin & Rudenko 2013).

As for the process of the agricultural sector informatization, it is worse than in Ukraine in the whole. It is explained by the peculiarities of the agro-industrial complex. Agriculture is an ideal environment for the use of modern information technology. However, the lack of funds in the area of agricultural science and production does not assure their widespread use (Tzyferova 2012).

Besides, the role of the state has not been determined in the agricultural sector in this area so far. A separate national program on informatization

and automation of agriculture is necessary. In this case, work steps, targets and results must be clearly defined. It is important to consider this issue in a complex way and it is necessary to design a system that would take into account a wide range of information agribusiness development (Tzyferova 2012).

Conducting an analysis on the implementation of ICT in agriculture, it can be concluded that compared with the other sectors of economy there is a noticeable lag in terms of their use. It can be explained by the following main points: the lack of facilities in the majority of modern computer technologies, unpreparedness or lack of qualified experts in information technology, the lack of appropriate information and software that allows you to automate the management of enterprises of agricultural sector (Tzyferova 2012).

Among all abovementioned reasons, the latter is the most important. It is explained by the fact that the purchase of computer equipment and the training of relevant professionals to use it do not cause any problems nowadays. With regard to the development and the creation of software and information management, there are some difficulties which are primarily associated with the lack of appropriate techniques that would allow to use computers and related software in the process of the fulfillment of the management tasks by agricultural production to the whole extent (Tzyferova 2012).

Despite a very large number of problems in the implementation and the provision of the agricultural sector with the latest information and communication technologies, in Ukraine, a constant work is being performed to improve the current situation.

First of all, the training of the qualified staff is assured for the information and analytical provision of the branches of Ukrainian economy. A contribution to the spread of new information technologies and the systems for information and communication software is made by the educational institutions that train specialists for the agricultural sector. It is noteworthy to draw attention to the National University of Life and Environmental Sciences of Ukraine (NULES), the largest agricultural university of Ukraine where more than 26 thousand students are studying. Besides the Information Technology Department, the University includes Ukrainian Education and Research Institute (ERI) of the information and telecommunication support of agricultural and environmental sectors of economy (Ukrainian ESI 2016).

For the purpose of students' practical training, specialists of the leading world manufacturers of hardware and software are involved. In particular, under the "IBM Academic Initiative" program, the IBM Corporation in Ukraine holds educational seminars and training courses at the Faculty of Computer Science and Economic Cybernetics. The software of the companies Apple, Micro Strategy,

Sun, IBM, SearchInform, etc. started to be implemented in the learning process.

The Research Institute for Information Technologies in Environmental Management, which is a part of ERI, combined the scientific activity of the research and innovative laboratories. During the first years of its functioning, the experts were conducting a study and completed a number of IT application lay-outs (Ukrainian ESI 2016).

Among the useful innovations that have been introduced and have already received a due recognition, there is an information-analytical system of monitoring the socio-economic development of the agro-industrial complex of Ukraine. This analytical software set of monitoring agricultural market objects is based on the technology that allows to conduct the structural analysis on the location and the infrastructure of agriculture as well as a list of characteristics to them. The results are represented by using the most informative map service. This lay-out was awarded the golden medal of the international exhibition "Agro-2012"; it allowed to develop an effective model of a system of storage, display and analysis of the socio-economic information on the status of agricultural and environmental sectors on the basis of the use of a software platform of the Microstrategy business-analytics (Ukrainian ESI 2016).

The software system of information and analytical support of the producers in the crop of spatial data was developed ("Agro-2012" Diploma); the geoportal of NULES of Ukraine with a set of the geological and informational services was created. A research was conducted in several areas within the state program on the development of information technologies for the agricultural sector of Ukraine.

A contribution of the Institute in the improvement of the access to agricultural information became the creation of the Internet portal "The Agricultural Sector of Ukraine" AgroUA.net (<http://agroua.net/>).

The purpose of its creation was the development of the universal and the comprehensive information resource to meet the needs of agricultural information, agricultural producers, commercial organizations, advisory services, researchers, teachers, students and other users (Ukrainian ESI 2016).

Recently, in the collaboration with leading advisory services and their leaders, the staff of the Institute is successfully creating a system of electronic counseling eXtension on the basis of the NULES of Ukraine.

It is of common knowledge that the lack of information provision in the modern Ukrainian village is the reason for numerous problems. Agricultural practice of the majority of developed countries points to the need for continuous training of producers, distribution of agricultural information and knowledge. Therefore, it is necessary to provide the powerful channels of knowledge dissemination. Nowadays

there are the qualitatively new forms of operational information dissemination based on the use of the advantages of modern computer networks and telecommunications. Special attention must be paid at the US experience in developing an electronic system of agricultural extension, which was called eXtension. The studied experience can be used for the creation of the electronic extension system (*e-Extension*) in Ukraine (NULES of Ukraine 2016).

It is necessary to underscore that during a long period of time, the system of providing information and consultation fulfills a difficult task of disseminating information and knowledge among agricultural producers of Ukraine. The purpose of this project was to found an electronic educational advisory system (*e-Extension*) for the informational support of farmers, inhabitants of rural areas and rural areas development electronic extension system (*e-Extension*) in Ukraine (NULES of Ukraine 2016).

The task of the electronic system of *e-Extension* extension is to provide objective, scientific-technical and educational information for the public to answer the users' questions. It is achieved by creating a national online database of high-quality information, which is based on innovation and the concept of sustainable development of agriculture electronic extension system (*e-Extension*) in Ukraine (NULES of Ukraine 2016).

The system developers have foreseen and defined the basic idea of the project that is the following:

- *e-Extension* must meet the needs of users who want to obtain the necessary information “in any place and at any time” more properly, give them a quick access to the resources of the organized personalized access that is necessary to make informed decisions;
- *e-Extension* applies to modern Internet technologies to help the counselors at the national level by providing the necessary information support to the users of agricultural goods and promoting the establishment and the development of the practitioners community by organizing discussion groups, establishing local contacts and interacting with the counselors of all existing services;
- information databases and *e-Extension* services must be available via the Internet to a wider audience of web users who can access the educational resources at any time in the various subject areas;
- *e-Extension* users will be able to find an objective, evidence-based information collected by universities, research centers and experts in the whole system of Advisory Services electronic extension system (*e-Extension*) in Ukraine (NULES of Ukraine 2016).

Newsletters, innovation, on-line responses, thematic discussion groups and training modules,

everything that is created by the advisory service specialists and the related industries, will help the users to find the information they need quickly.

All the abovementioned issues will make it possible to improve the state of information and consultation provision of the agricultural sector of our country and the communication capabilities of advisory services in the future. It is worth noting that over the last decade, advisory service made a significant contribution to the development and the establishment of an information system of the agricultural sector of Ukraine.

In Ukraine, the process of creating advisory services became the most prevalent after the adoption of the Law of Ukraine “On Agricultural Advisory Activities” of 17 June 2004. According to the latest register of the Ministry of Agrarian Policy of Ukraine, in the beginning of 2016, 71 Agricultural Advisory Service was founded in Ukraine (in 2008 there were only 26 of them) (Chekhlatyi 2008).

In Poltava region, the first advisory service called “Poltava Regional Agricultural Advisory Service” (PRAAS) was established in 2004. This project was a result of the intergovernmental agreement of 15 June 2004, signed between the Federal Ministry of Food, Agriculture and Consumer Protection and the Ministry of Agrarian Policy of Ukraine. Much effort was put by the teachers of Poltava State Agrarian Academy to found PRAAS (Chekhlatyi 2008).

Along with the establishment of advisory services, information technologies such as Internet resources were involved to advise producers of agricultural workers and leading specialists of the field in order to improve the efficiency of their operations. The site “Tip”, which is a universal resource to meet the information needs of the agricultural areas, was designed. The site has a large number of structured sections which represent an abundance of materials on the most pressing issues for agricultural producers. Among them, the most significant ones are the following sections: Accounting, Current Issues, Beautification, Livestock, Crops, News, Tips and others. On the advisory service page there are the links to the sites that are the most popular with farmers as well as a forum for the public discussion of important issues and problems. It is also possible to write a personal message, send a request to the administrator or ask questions relating to the scope of the service moderator. On the site, a lot of practical advices on agricultural advisory activities, manufacturing issues that a counselor faces with in everyday work, news and posts that directly relate to agricultural producers can be found. Here the experience and recommendations by the leading experts in the agricultural area, Poltava State Agricultural Academy (PSAA) and the Department of Agricultural Development of the Poltava Regional State Administration were gathered (Chekhlatyi 2008).

Besides the web-site of Poltava regional agricultural advisory service, its employees have their personal Internet blogs created on the free service Google—Blogger.com (blogspot), where they are able to share the information on the issues that are the most interesting for agrosphere.

Due to the increase in the demand for qualified advisory services, the need to solve the lack of effective mechanisms for the agricultural science cooperation, education and agriculture, was faced in the foundation of advisory services which would be alternative to the existing ones in Poltava region. Applying to the existing technical, scientific and organizational potential of PSAA it could provide effective social-focused advisory services to agricultural producers and the rural areas.

This service was established in 2007 on the basis of PSAA, and it was given the right to provide the socially oriented advisory services using state budget. This body was called Poltava Regional Public Organization “Official Agricultural Advisory Service” (TRPO “OAAS”) (Kalinichenko et al. 2011).

In order to disseminate the information on the activities of a new advisory service an Internet resource was established; it was the site called TRPO “OAAS”. However, the financial crisis which took place in Ukraine in 2008 had a negative effect on financing the service activity and the design as well as the support of its site. Taking into consideration that the work of the service was not organized on a commercial basis, but was aimed to provide the socially oriented advisory services from the state budget, a sharp reduction of state funding in almost ten times in 2009 in comparison with the previous year, raised its activities in an extremely difficult situation. However, the “Official Agricultural Extension Service” has not slowed down its active functioning despite the difficult financial situation of its funding from the state. Its work has shifted largely to cooperation with the international organizations and projects. Besides, thanks to the cooperation with the local authorities, another source of funding became the local budget (Kalinichenko et al. 2011).

TRPO “OAAS” has gained a considerable experience in the international cooperation and participation in various projects, which makes it possible not only to adopt the best practices, but also the necessary funds for a variety of educational events. One of the results of such projects was the creation of a training manual for the distance learning in advisory services (Kalinichenko et al. 2009).

Using modern information technologies, which enable to carry out distance learning without a direct personal contact between a teacher and a learner, was laid on the basis of distance education. An important and a necessary part of distance education are electronic manuals (Pustovit 2008). In order to train the specialists of advisory

services and the students of the higher educational establishments, the scholars of PSAA developed an electronic guidebook in advisory services which is correspondent to the special curriculum and was ordered by the Ministry of Agrarian Policy and Food of Ukraine; it included 5 modules: Theoretical, Legal, Economic, Ecological training and ICT. Besides, the guidebook provides the necessary background information and the modules are accompanied by the tests for self-control (Pustovit 2008).

This tutorial is the first electronic textbook in Ukraine, which is classified by the Ministry of Agrarian Policy and Food of Ukraine.

In the process of developing an electronic handbook, programming languages that are modern and available for mastering have been used, namely, HTML, XML, JavaScript. To ensure software compactness program wrapper was used.

It should be highlighted that electronic textbooks and manuals have a number of advantages. They include the automation of data storage; virtually unlimited amount of information, relatively low production costs. It is necessary to notify that structuredness, user-friendliness and clarity of the material in the manual are assured by using hypertext. A user has an opportunity not only to “surf” the pages, but also he/she can manage the issuance and the acceptance of the material (Pustovit 2008).

PSAA and TRPO “OAAS” successfully completed another joint ecological project “Tempus-Tacis 2006” (JER_27168_2006) which was aimed to found the “Agro-Ecological Center of Poltava region”; it had to address such major challenges in the environmental areas as:

- improvement of the environmental education;
- gaining experience by the Ukrainian experts in the area of environmental problems;
- study of the possibility for the introduction of the international standards for environmental protection in Ukraine;
- dissemination of environmental information and results of studies [Kalinichenko & Chekhlatyi 2009].

In order to disseminate information about the work of the Centre, the site “International Agro-Ecological Center” was designed. For agricultural producers, both in a certain region and Ukraine as a whole, the information posted on the website of the Centre is very valuable. In the issue of effective distance learning of agricultural producers, an idea of a digital library center is interesting. The site contains the electronic publications of the series “Environmental Library of Poltava region”, booklets on the state of environment in Poltava region, which provided analytical information on air quality, water resources, drinking water quality, land resources, waste and recreational resources of Poltava. Electronic ecological library of the center continues to be improved and updated with

new material environmental issues. Currently, it contains hundreds of books, manuals, textbooks, scientific papers and various materials on agroecological subject already [Kalinichenko et al. 2009].

It is evident that the main task of advisory services in the agricultural sector of Ukraine is to disseminate information among agricultural producers. Their efficient functioning is possible primarily due to the use of modern information and communication technologies by the agricultural manufacturers. The use of information technology in advisory services significantly reduces management costs, expands the access of the agricultural producers and rural populations to information sources and communications and facilitates profitable farming.

4 CONCLUSIONS AND SUGGESTIONS

Nowadays due to the the absolute cost, Ukraine occupies the last place in Europe in all indicators of information (density coverage area and capacity of telecommunication and computer networks, Internet users, the proportion of broadband internet to the total number, etc.). In the area of the access to the Internet, Ukraine considerably lags behind developed countries.

The experience of developed countries demonstrates that the use of the latest informational technologies and information support systems is a prerequisite for the high-tech agricultural production and management. It is of common knowledge that Ukraine is one of the world's largest potential agricultural producers. Improving information and communication for the agricultural sector provides a significant opportunity to increase the production of agricultural products significantly and to become one of the largest food manufacturers. Of course, these tasks fulfillment should help the producers and the state. Its aid should be one of the main priorities of the State agricultural and information policy. Considering the fact that advisory services make a significant contribution to the agricultural producers awareness, expanding their competitiveness, solving the problem of employment of the rural population there is a need to support them at the state level.

Currently, the most serious challenge to the existence of agricultural advisory services in Ukraine is primarily the practical lack of state funding. Nowadays the only way to solve this problem is to attract local budget funds, international grants and financial assistance to foreign investors. The situation may change for the better only if the financial crisis is overcome, and the economic situation in the country is improved. Only under improving the financial security of advisory services in Ukraine it will be possible to state on the prospects of

providing them with innovative technologies. In its turn, it will improve the efficiency of the work of agricultural advisory services, the quality of information and consultative support, promote the organization of competitive production.

REFERENCES

- Chekhatyi O.M., 2008. The Development and the Establishment of Agricultural Advisory Services in Poltava Region. *Scientific Bulletin of S. Z. Gzhytsky National University of Veterinary Medicine and Biotechnology in Lviv*, 489–493.
- Draft concept of the Design of an Advisory Services Electronic System (*e-Extension*) in NULES of Ukraine. Access: <http://edorada.org/> (the date of the appeal 23.3.2016).—Name from the screen.
- Information Society, Kyiv: *National Commission for State Regulation of Communication and Informatization*. Access: <http://www.nkrzi.gov.ua> (the date of the appeal 25.3.2016).—Name from the screen.
- Kalinichenko A., Chekhatyi O., 2009. Development and International Cooperation of the Informational Services and Consultation to Ensure the AIC in Poltava Region. *Bulletin of KhAI. Series "Economics of Agriculture and Natural Resources"*. Kharkiv: 60–68.
- Kalinichenko A., Chekhatyi O., Gorb O., 2011. The Formation and the Development of the Advisory Services System at Poltava State Agrarian Academy. *Collection of scientific and methodical works "Science and technique"*. Kyiv: Agricultural Education, 23: 52–58.
- Kalinichenko A., Chekhatyi O., Kostohlod K., 2009. Features of Information Support of Agricultural Enterprises in Poltava Region. *Proceedings of TSAU*, Melitopol: 354–361.
- Khudyakov H.O., 2016. The Role of Information Security in the Management of Agricultural Enterprises. *International Internet-conference "Formation and Development of the Economy under Current Economic Conditions"*. Access: <http://www.wp.viem.edu.ua/> (the date of the appeal 02.25.2016).—Name from the screen.
- Larin N.B., Rudenko O.M., 2013. Information and Communication Provision of Efficient Functioning of Power: *Educational and methodological materials*. Kiev: NAPA: 5–7.
- Pustovit S.V., 2008. International Project to Found an Agro-Ecological Centre in Poltava: *Materilas of the International scientific and practical conference "Methods of teaching natural sciences at high school"* (Poltava, 29–30 May 2008). MES Ukraine, Korolenko PSAU. Poltava: Astraya: 363–365.
- Tzyferova N.H., 2012. Information management mechanism of state regulation of a agriculture complex. *Scientific Notes National University "Ostroh Academy"*. Series: *Culture and social communication*, 3: 72–78.
- Ukrainian ESI Information and Telecommunication Support of the Agricultural and the Environmental Sectors of Economy. Access: <http://nubip.edu.ua/node/9488> (the date of the appeal 22.3.2016).—Name from the screen.

Vendor Managed Inventory—implementation of VMI concept from the dynamic management perspective

H. Kościelniak & M. Starostka-Patyk

Faculty of Management, Czestochowa University of Technology, Czestochowa, Poland

ABSTRACT: Currently constructed management systems are characterized by their activity directions that allow the processing of vision, mission, and strategy into operational objectives for particular groups of businesses entities or individual employees. This also applies to the modernization and restructuring of the supply chain. In the conditions of dynamic environment business entities are taking actions to strengthen the brand among suppliers, buyers, distributors, retailers and the local community. One of such actions area is VMI system. Moreover, efforts to implement Vendor Managed Inventory (VIM) correspond to the concept of dynamic management. This paper states as a part of lively discussion that goes on in the world literature on the growing importance of VMI system and the role of individual companies in its development.

The aim of this paper is the assessment of VMI adaptation and adjustment to the conditions of enterprises operation in the changing environment with connection to the concept of dynamic management. Literature studies in the field of supply chain management has been deepened by own research conducted among Polish enterprises implementing VMI. The value of this paper is releasing the tendencies in the VMI systems development, which is to support the process of dynamic management with the creation of synthetic knowledge about the enterprise and its surroundings, along with the directions of their updating and improvement.

1 INTRODUCTION

Today, in a dynamic environment, the sale of the product is more difficult than its production. It encourages enterprises to pay particular attention to the method of product distribution to the customer.

The essence of the distribution is to transfer products from the places of their generation to final customers at the place and time where they are awaited by buyers. The mistake is only the production and offering products at competitive prices. To succeed, an enterprise must deliver products in a proper way to the demanding market. From the market-oriented enterprise management point of view the distribution creates opportunities to build competitive advantage. Distribution functions arise from the necessity of eliminating the existing discrepancies between supply and demand, as well as differences in terms of type, quantity and range of products, and also differences of the time and place of their production, and in the final stage—consumption (Kadłubek 2012).

Distribution in terms of logistics is an integrated structure of the flow of products and accompanying information flows. The product flow includes all activities related to commodities, raw materials or final products and their storage, transportation,

packaging and orders handling. The sphere of information creates activities such as: order preparation, administrative developments and launching of the goods flow. All these elements can be summed up in so-called distribution network. It has a spatial character and a certain order of the territorial division on the distribution of manufacturers and warehouses together with links to adequate transport infrastructure.

Optimal distribution network is the result of the strategies and development plans analysis for the company and logistic parameters and costs in the current network. Analysis of the existing logistics system, distribution network, costs analysis and calculations, analysis and forecast of goods flows and analysis of the company development plans allows to find the right configuration of the distribution network, taking into account logistics and business security of enterprise, as well as corresponding plans for its development.

Recently, two generic strategies for supply chain design have emerged: efficiency and responsiveness. Efficiency aims to reduce operational costs; responsiveness, on the other hand, is designed to react quickly to satisfy customer demands. A crucial question in the supply chain is the design of distribution networks and the identification of facility locations. Ballou & Masters (1993) put

forward four strategic planning areas in the design of a distribution network system, as shown in Fig. 1.

The first issue deals with customer service levels. The second one deals the placement of facilities and demand assignments made to them. The third deals with inventory decisions and policies that involve inventory control. The fourth deals with transportation decisions of how transport modes are selected, utilized, and controlled. All four of these areas are inter-related and the customer service level is determined by the other three decision areas. There are practical challenges for firms when they try to simultaneously reduce operating costs (for efficiency) and customer service (for responsiveness). In traditional supply chain network design, the optimization focus is often placed on minimizing cost and maximizing profit as a single objective. However, very few distribution network systems should be considered as intrinsically single objective problems. It is not always desirable to reduce costs if this results in a degraded level of customer service. Thus, it is necessary to set up a multi-objective network design problem (Liao et al. 2011).

Planning a distribution network is based on the need for:

- cost optimization;
- changes in the scope of action;
- assessment of the logistics system safety;
- (re)locate warehouses;
- improving the logistic parameters;
- better use of transport.

Properly developed distribution network matching the enterprise profile provides, inter alia, the possibility of obtaining benefits such as: optimizing the cost efficiency of distribution, securing timely deliveries, strategic security in logistics, increasing the competitiveness of enterprises, the elimination of inefficiencies in the current functioning of the distribution network, shortening delivery time and optimizing the level of stocks.

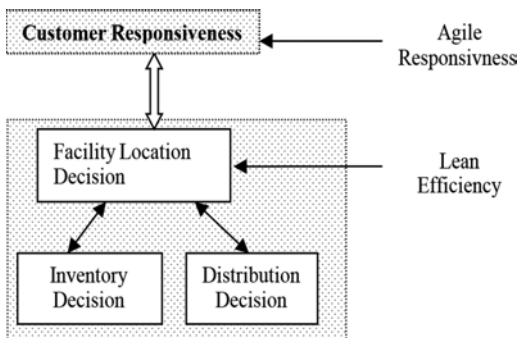


Figure 1. Four strategic planning issues in distribution network design.

2 VENDOR MANAGED INVENTORY ESSENCE IN DYNAMIC ENVIRONMENT

The conditions in which enterprises operate today make that many of the existing concepts of management do not help managers manage their companies according to internal and external changes (Lan 2009; Brzozowska et al. 2015). Managers must still undertake the tasks of adapting enterprise behaviour to the existing reality. The management system should allow the company to identify and analyse the changes and for continuous adjustments of the methods and actions, also in the scope of supplies.

Vendor Managed Inventory (VMI) is one of the tools of Supply Chain Management (SCM) used to deliver the right product in the right place at the right time and in the right quantities.

The concept of Vendor Managed Inventory in practice appeared for the first time in the '80 s in the retail industry, implemented by the two partners—Wal-Mart and Procter & Gamble. However it is indicated that the issue of responsibility for inventory management was already taken in the literature in 1958 by John F. Magee.

The essence of the VMI concept is the transfer of competences and the related responsibility in the supply chain. The cell located closer to the end consumer in the supply chain resigns from one of its competences (inventory management) for a partner—its supplier. Potential benefits of the VMI concept came from the assumptions for its justification: that the supplier (as the practice shows—usually the manufacturer of specified range products) has a better knowledge of its own products, especially their manufacturing process (e.g. the time needed for their production); and that the supplier is responsible for supplying the recipient only in a limited number of its own products, so will do it more effectively than distributor operating at much wider products range. Reducing inventory levels throughout the supply chain have to be the result of a deeper chain penetration by the information on demand and enhanced partners cooperation. In turn, a better knowledge of the products life cycle on the market, resulting from more accurate knowledge the demand by manufacturer—would avoid the unpleasant consequences of obsolete products dispose that fill up their warehouses.

VMI is a system, a form of cooperation between the supplier and the recipient of trade, which involves reversal of the responsibility for replenishment the supplies to the supplier, in return for which the recipient (for example a given retail network) shares accurate information about the current state of storage and the current descents from the store (Lambert, & Schwieterman 2012). VMI

means to optimize the functioning of the supply chain as a result of recipient inventory management by the manufacturer, who determines the time and volume of delivery, while ensuring full availability of products.

In a vertically decentralized supply chain, the Vendor-Managed Inventory (VMI) system with revenue-sharing consignment contract streamlines decision-making processes and flows of goods and information. This business model is widely used in various industries, including personal computers, sports equipment, automobiles, clothing, furniture, firearms, music, tools and antiques (Chen 2013). VMI is also becoming an important element of modern supply chains. They are created by the idea of the so-called supply chain 2.0 and their primary purpose is to achieve structural flexibility; exemplification of this chain on the example of Hewlett Packard and World Duty Free has been shown by Christopher & Holweg (2011) and Szymczak & Grabański (2013).

3 THE POTENTIAL TO CREATE VALUE THROUGH THE IMPLEMENTATION OF VMI

One of the objectives of VMI is the reduction of inventory levels, which is consistent with the goal of efficient supply chain, minimizing costs of efficient supply chain with the right product.

Another of the objectives of VMI is a pull the flow of material, according to the data of current demand—which corresponds to the purpose of responsive supply chain (moving towards the convergence of supply and demand responsive supply chain suitable for innovative products).

Its smooth operation significantly reduces the time of stocks storage, which is equivalent to reduction in costs for both, the one and the other; also allows the reduction of gaps in inventories, which improves the level of customer service (Yao et al. 2010).

The supplier has full authority over inventory management at the buyer's DC to pay all costs associated with the supplier's production cost, both the buyer's and the supplier's ordering cost, the inventory holding cost and distribution cost. The supplier monitors, manages and replenishes the inventory of the buyer. Thus, the decisions on order replenishment quantity and order shipping are given to the supplier in the VMI system, rather than to the buyer as in tradition systems. Fig. 2 presents the operational cost structure between the partners in the VMI system (Liao et al. 2011).

VMI is part of a strategic, process and functional levels of dynamic management. At the strategic level, area of enterprise development

decisions VMI is a key field of the main decision-makers and the company board activities. For the process level the managers' mid-level is responsible; it is a collection of functions assigned to the appropriate departments. For the functional level other employees are responsible who implement appropriate actions assigned to them.

The implementation of VMI requires recognition of the advantages and disadvantages for both, the suppliers and recipients (See Tab. 1, 2, 3, based on Szymczak & Grabański 2011; Franaszek & Sadowski 2008).

Prior to the reduction of costs and reduction of risk the enterprises strengthen their cooperation in order to enhance the ability to compete or combine complementary skills and acquire knowledge in order to acquire new competencies. Partnerships require trust here, communication and exchange of knowledge (Światowiec-Szczepańska 2012; Hamel 2006; Bubel 2015). Taken in the supply chain and in VMI collaboration represents the excess of the benefits resulting from its adoption in relation to activities carried out by the company itself.

Model of value creation path, from the factors of cooperation to the financial results is presented on Figure 3 (Simatupang & Sridharan 2005; Nowicka 2011).

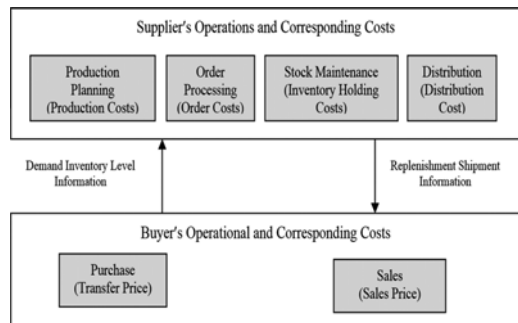


Figure 2. Cost structure of VMI system.

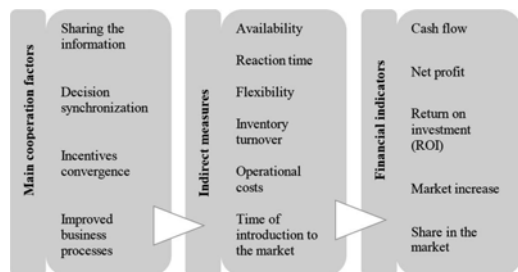


Figure 3. Model of value creation path, from the factors of cooperation to the financial results.

Table 1. Advantages of VMI implementation.

Advantages for supplier	Advantages for recipient
Realization of integrated supply chain concept;	Realization of integrated supply chain concept;
Reduction of inventory levels;	Reduction of inventory levels;
Lower level of capital tied up in stocks;	Lower level of capital tied up in stocks;
Unified way of communication between partners;	Unified way of communication between partners;
Automatization of receiving and handling orders process;	Automatization of ordering process;
Higher level of relationship with the recipients;	Reduction of the role and relieving the purchasing department;
Transparency of recipient and its stocks demand;	Elimination of gaps;
Optimization of production batches;	Minimization of errors in the ordering process;
Increase the number of on-time delivery;	Possibility of negotiating the additional discount;
Faster response to changing demand;	Supplier faster response to changing demand;
Optimization of recipient supplies process.	Opportunity to focus on marketing and sales.

Table 2. Disadvantages of VMI implementation.

Disadvantages for supplier	Disadvantages for recipient
High implementation costs;	High implementation costs;
Responsibility for recipient inventory;	Disclosure to supplier the sales data;
Increased demand for analytical services;	The increase in the degree of dependence on suppliers;
The need to improve algorithms for estimating demand;	Risk of errors made by the supplier;
The need for more flexible production and transport processes.	Increase the role of the controlling department.

Table 3. Benefits of VMI implementation for supplier and recipient.

Beneficial area	Change value	All benefits
Benefits for supplier		
Lower level of stocks	30	Reduction of distortions in demand and the improvement of its transparency.
Lower costs of transport	10	Focusing on the demand of the final recipient; Reduction of transport costs.
Lower costs of warehousing	13	Reduction of administrative costs; Optimizing the size of production runs.
Shorter delivery time	50	Errors reduction.
Better customer service	>10	Increase recipient loyalty; Creation the way to handle the logistics that will be difficult to imitate by competitors; Obtaining a closer relationship with recipient (the possibility of introduction more advanced programs based on partnership in the supply chain, for example, joint management of product category).
Benefits for recipient		
Lower level of stocks	10	Reduction of inventory levels through the direct impact on stocks.
Higher sales level	8–10	Reduction of administrative costs.
Lower logistics costs	3–4	Increase of products availability Reduction of supply cycles.

The process of creating value includes the qualitative and quantitative measures, which examine the logistics steps, also within VMI.

4 THE ASSESSMENT OF VMI IMPLEMENTATION IN MANUFACTURING ENTERPRISES OF METAL INDUSTRY

In Polish manufacturing enterprises of the metal industry (small and medium-sized) was conducted an own research on the VMI implementation. Currently on the Polish market are more than 53

thousand of companies in the metal industry (production of metals and metal products). The largest group (about 91%) are companies employing up to 9 workers and from 10 to 49 workers (about 75%).

The main problem of this industry is large fragmentation and the share of small/medium-sized companies. It should be noted that last year, many companies ended their business. In 2013, the bankruptcies were recorded 886 times, so 1% more than in 2012. Problems in the metal industry are mainly due to its close links with other sectors, including construction, industrial engineering and automotive, as well as with diversified relationship in the supply chain. Firms also faced with a surplus of production capacity, which are up to 30–40%. But there are also companies that positively assess their financial situation.

The pilot research was carried out in January 2016 year. The number of sent by e-mail questionnaires were 112. Survey questions were answered by 76 people (managers of top and middle level). Empirical studies were related to:

- Recipients satisfaction on the VMI implementation;
- Recipients dissatisfaction on the VMI implementation;
- Delays on the VMI implementation;
- Assessment the market position of enterprises during the VMI implementation.

Figure 4 shows the opinions on the implementation of VMI, taking into account the time of its implementation. The biggest number of VMI implementation was completed within 6–12 months (37% of responses); another group constitute that the VMI implementation was covering over one year (32% of responses) and 6 months (29% of responses). The time of VMI implementation was indicated by 29% of respondents (see Figure 4).

Figure 5 summarizes the reasons for managers' dissatisfaction on VMI implementation. The most common reason for dissatisfaction was the failure to meet the expectations of the functional system (24% of responses). This may result either from improperly planned range of VIM, or the complexity of the processes within the company and the high costs associated with the modification of the system; it could cause the resignation of the system adaptation to the needs of the company (see Figure 5 and also Figure 2).

From the empirical research comes out that an important issue in the realization of the functional expectations in connection with the VIM implementation is estimating the demand. In VMI algorithm the demand is estimated not only on the basis of historical data, but also is taken into account the current market situation (Park et al. 2016; Sadeghi et al. 2015). In determining the algorithm to guarantee the

profitability of production, into account are taken the costs of retooling machines, other production costs and deliveries the goods to the customers; also the size of the minimum production at the acquired margins is calculated. VIM provides the basis to shorten the planning horizon to the minimum optimum length, and hence, to keep the time to react, when the earlier forecast of demand proves to be misguided. The VMI implementation has many pitfalls of an informative character, which can reduce the benefits. Detailed description of the general dissatisfaction is insufficient trust and communication between partners. The surveyed companies proved that it is about lack of the continuity and deepening of knowledge sharing in partnerships for the VMI implementation. Moreover, managers are unable to identify solutions to improve this scope of cooperation in the supply chain.

In addition, the respondents pointed out also the most frequently cited errors in the process of VIM implementation (see Figure 6).

Among the three most frequently cited errors on VMI implementation the following belong in the order: incomplete knowledge of demand (34% of responses), low organizational cohesion (31%) and erroneous determination of the VIM implementation (23% of responses). Errors in the field of

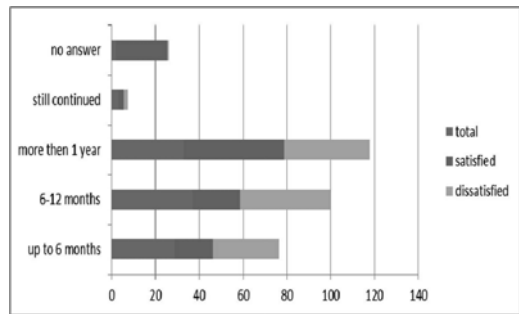


Figure 4. The level of recipients' satisfaction on VMI implementation.

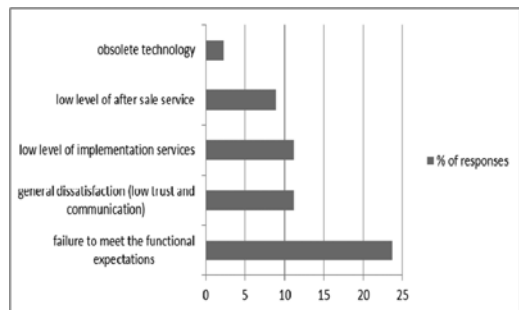


Figure 5. The reasons of dissatisfaction on VMI implementation.

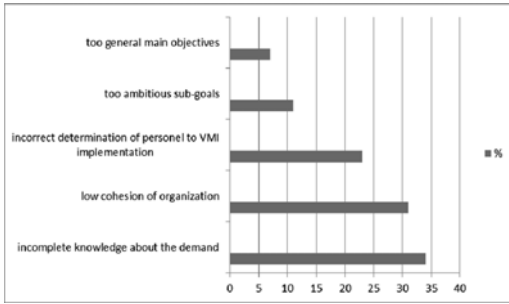


Figure 6. Errors in the process of VIM implementation.

Table 4. The effects of VMI implementation on enterprise market position.

	How the VMI implementation influenced the enterprise market position?	% of responses
1	Increased market position	71
2	No change in market position	14
3	Decreased market position	5

over-ambitious partial and general objectives are combined with the problem of estimating demand. The respondents indicated that in majority they estimated the main objective on the basis of quantitative criteria, and only 25% of qualitative criteria. This could lead to the designation of too ambitious main or sub-targets, thereby limiting the effectiveness of the entire process of VMI.

The implementation of VMI was verified among managers of surveyed companies by determining its effect on the market position (see Table 4).

The overwhelming majority of respondents (71%) indicated an increase in its market position as a result of the VMI implementation. The respondents indicated that properly implemented VMI contributed to the higher sales (62%) and improve the economic and financial results (66%). This is a very important area of research because managers recognized the increased competition as the most important determinants of dynamic market changes.

5 CONCLUSIONS

The paper presents an attempt to evaluate the VMI adaptation and adjustment to the conditions of operation in a changing and dynamic business environment. The study suggests that the advantages and benefits of VMI system for the supplier and recipient correspond with the characteristics of dynamic management.

The use of VMI system requires (Platonoff 2009):

- Continuous adaptation to changes inside and outside the environment of supplier and recipient;
- Continuous learning of employees and improve the collection of knowledge about the data on supply chain;
- Complexity look at supply chain management through the involvement of all resources, tangible and intangible of partners which reflects the characteristics of the concept of dynamic management.

What is more, VMI illustrates how inventory management is useful for managers of enterprises and gives the ability to meet the expectations of all stakeholders (in short-, medium- and long-term) involved in the supply chain. These are further features of the concept of dynamic management implemented by VMI.

Benefits of VMI are also appreciated by Polish enterprises implementing VMI system. Their satisfaction grows with the lengthening of the period of VMI application. The tangible result of the use of VMI is determined improvement their market position. However, Polish companies should make further efforts to correct the mistakes and difficulties in the process of VMI implementation and fully utilize and adapt the possibilities offered by VMI for partners in a dynamic environment of enterprises chain management.

REFERENCES

- Ballou, R.H., & Masters, J.M. 1993. Commercial software for locating warehoused and other facilities. *Journal of Business Logistics*, 14(2): 70–107.
- Brzozowska A., Grabińska A. & Dacko M., 2015. Evolution of Supply Chain Management and Striving to Achieve Sustainable Development, *Logistyka* 3: 24–30.
- Bubel D., 2015. Risk Management in Logistic Projects, *Carpathian Logistics Congress (CLC 2015)*, Jesenik, Czechy (04 do 06 listopada 2015 r.), Tanger: 32–33.
- Chen, L.T., 2013. Dynamic supply chain coordination under consignment and vendor-managed inventory in retailer-centric B2B electronic markets, *Industrial Marketing Management*, 42: 518–53.
- Christopher M. & Holweg M., 2011. Supply Chain 2.0. Managing Supply chains in the Era of Turbulence, *International Journal of Physical Distribution & Logistics Management*, Vol. 41, No. 1: 63–82.
- Franaszek J. & Sadowski A., 2008. Zastosowanie Vendor Managed Inventory w Kujawskiej Fabryce Manometrów “KFM” SA, *Logistyka*, no 1: 34–37.
- Hamel G. 2006. Alianse strategiczne. Sztuka zdobywania korzyści poprzez współpracę, Helion, Gliwice.

- Kadłubek M., 2012. Zarządzanie procesami dystrybucji w przedsiębiorstwie, cz. 1, *Logistyka* no 5.
- Lambert D.M. & Schwieterman M.A., 2012. Supplier Relationship Management as a Macro Business Process, *Supply Chain Management: An International Journal*, Vol. 17: 341.
- Lan H., 2009. Corporate strategic management: Static and dynamic paradigms, *Research Article, Front. Bus. Res. China*, 3(1): 50–62.
- Liao S.H., Hsieh Ch.L. & Lai P.L., 2011. An evolutionary approach for multi-objective optimization of the integrated location inventory distribution network problem in vendor-managed inventory, Elsevier, *Expert Systems with Applications* 38: 6768–6776.
- Nowicka K., 2011. Współpraca partnerska w łańcuchu dostaw, *Gospodarka Materiałowa i Logistyka* no 6.
- Park Y.B., Yoo J.S. & Park H.S., 2016. A genetic algorithm for the vendor-managed inventory routing problem with lost sales, *Expert Systems With Applications*, Elsevier, <http://dx.doi.org/10.1016/j.eswa.2016.01.041>.
- Platonoff A.L., 2009. Zarządzanie dynamiczne. Nowe podejście do zarządzania przedsiębiorstwem, *Difin*: 40–41.
- Sadeghi J., Taghi S. & Niaki A., 2015. Two parameter tuned multi-objective evolutionary algorithms for a bi-objective vendor managed inventory model with trapezoidal fuzzy demand, *Applied Soft Computing* 30: 567–576.
- Simatupang T.M. & Sridharan R. 2005. Supply chain discontent, *Business Process Management Journal*, vol. 11 No. 4.
- Światowiec-Szczepańska J., 2012. Ryzyko partnerstwa Strategicznego przedsiębiorstw. Ujęcie modelowe, *Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu, Poznań*.
- Szymczak M. & Grabański Sz., 2013. Vendor Managed Inventory—właściwe podejście do zarządzania łańcuchem dostaw w czasie kryzysu, *Gospodarka Materiałowa i Logistyka* nr 1:8.
- Yao Y., Yan D. & Dresner M., 2010. Managing supply chain backorders under vendor managed inventory: An incentive approach and empirical analysis, *European Journal of Operational Research* 203: 350–359.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Logistics chain management elements at global market of Liquefied Natural Gas (LNG)

M. Zawada & M. Starostka-Patyk

Faculty of Management, Czestochowa University of Technology, Czestochowa, Poland

ABSTRACT: Natural gas is currently at the world the most desirable source of energy. This is not only the fact of its low prices, which in recent years have steadily decreased, but primarily from an increase in its availability. Not without significance is the impact on the environment. Natural gas is less harmful to the environment than other energy sources: the emission of CO₂ from the combustion gas is 30% lower than in the case of oil and up to 60% less than that of carbon. It is also significantly reduced the emissions of other chemicals, including mercury, sulfur, and nitrogen dioxide. An important element is also its importance in ensuring the energy security of importing countries the blue fuel.

The aim of this study is to characterize the global LNG market for the years 1970–2014, taking into account all the links in the logistics chain of LNG and its management elements such as transport and distribution, from the place of gas extraction until its delivery to the final consumer.

1 INTRODUCTION

According to the published in February 2016 the latest report of BP (BP Energy Outlook Edition 2016), the global energy demand will increase in the years 2014 to 2035 by 34% at a rate of 1.4% per year. Crude oil and natural gas will continue to be the main fuels that meet the growing demand for energy. Demand for natural gas will grow the fastest among fossil fuels at a rate of 1.8% per year, and demand for crude oil will increase steadily by 0.9% per year, although its share in the energy mix will continue to decrease.

The clear increase in the demand for natural gas in recent decades all over the world, mainly comes from the possibility of its use in many sectors of the economy, both for the industry, the service sector and households. In many countries, natural gas is also used in the production of electricity. Taking into account the proven reserves of energy and their impact on the environment, natural gas is the primary energy source of the world, as a fuel with the least negative impact on the environment.

Natural gas is a fossil fuel of organic origin. Its properties set it apart, however, from other conventional sources of energy. The first essential advantage of the natural gas is relatively low extent of environmental pollution associated with its combustion. Emissions of carbon dioxide is almost less than half the emissions resulting from the combustion of coal or lignite. Secondly, the emission of substances that cause smog compared with other fossil fuels is also from 60 to 90% lower. It all makes, that natural gas is the cleanest fossil fuel on the market.

Consumption of natural gas in the world shows a steady upward trend. In the last ten years (2005–2014) the average annual growth was 2.3%, which resulted in an increase in annual consumption of natural gas by 617 billion cubic meters. The only region which consumes less natural gas than ten years ago is the European Union.

One of the fastest growing sub-sectors of the energy sector in the world is the production of Liquefied Natural Gas (LNG) and its trade. Expected global growth in demand for LNG is 10–12% per year. The development of LNG technology and LNG markets makes that regional gas trade is beginning to have a global character. LNG as a fuel of the future requires the development of adequate infrastructure. The efficiency of the LNG logistics supply chain functioning is conditioned not only by the presence of gas terminals, an essential link is also suitable fleet of ships for the transport. The aim of this study is to characterize the global LNG market for the years 1970–2014, taking into account elements of logistics management in the LNG supply chain, especially distribution and transport, from the point of gas extraction until its delivery to the recipient.

2 LIQUEFIED NATURAL GAS (LNG) CHARACTERISTICS

Liquefied Natural Gas (LNG) is obtained in the liquefaction of natural gas during which it is cooled to a temperature of -163°C . Its volume is reduced during this process with about 600 times. Changing

its physical state allows for efficient storage and transport. LNG as a liquid is neither explosive, toxic, nor corrosive. Evaporates under atmospheric pressure at temperatures above minus 162°C. Methane, the main component of natural gas becomes explosive when mixed with air at its share of 5% to 15% of the mixture. Ignition temperature is about 540°C (for a mixture of 10% methane in air).

The process of natural gas liquefaction is associated with the very precise purification of carbon dioxide, nitrogen, propane—butane, moisture, helium and other impurities. Liquefied Natural Gas (LNG) is a colorless, odorless, has no corrosive properties and is not toxic. Energy gas liquefaction with the purpose to their transportation is used for many years due to the environment friendliness, the ability to cover peak demands for gas and the stability and security of supplies.

3 GLOBAL MARKET OF LIQUEFIED NATURAL GAS (LNG)

The global market of LNG is growing at a very fast speed (Figure 1). In the period of 44 years (1970–2014) it has increased from 3 billion m³ to the level of 333 billion m³ (which is the increase in absolute terms by 330 billion m³, and in relative terms by 11 000%!). The average annual growth rate of LNG trade is 44%. In the last decade (2004–2014) this market has almost doubled.

The beginning of international LNG trade is considered the year of 1965, when the first route of liquefied natural gas supplies has been opened from the terminal Arzew in Algeria to Canvey Island in the UK. The leading exporters of LNG, with a state from the year 2014, (Table 1), include: Qatar with 31.9% market share, Malaysia which has a 10.5% of market share as well as Australia and Indonesia with shares respectively 9.7% and 8%.

Twenty-nine countries imported LNG from the global market in 2014. Europe had the world's

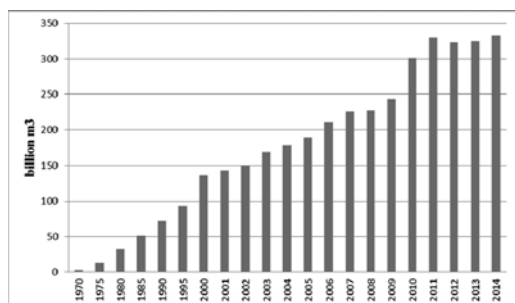


Figure 1. Global LNG trade volume from 1970 to 2014 in billion m³ (www.statista.com 2016).

Table 1. Leading exporters of LNG in 2014 in MTPA (World LNG Report 2015).

Country	Export
Qatar	76,8
Malaysia	25,1
Australia	23,3
Nigeria	19,4
Indonesia	16,0
Trinidad	14,4
Algeria	12,8
Russia	10,6
Oman	7,9
Yemen	6,8
Brunei	6,2
UAE	5,8
Peru	4,3
Eq. Guinea	3,7
Norway	3,6
PNG	3,5
Angola	0,3
Egypt	0,3
USA	0,3

Table 2. Leading importers of LNG in 2014 in MTPA (World LNG Report 2015).

Country	Import
Japan	88,9
South Korea	38,0
China	20,0
India	14,6
Taiwan	13,6
UK	8,5
Spain	8,2
Mexico	6,9
Brazil	5,7
Turkey	5,4
Argentina	4,7
France	3,3
Chile	2,8
Kuwait	2,7
Singapore	1,8
Malaysia	1,8
Other	9,5

only new importer, Lithuania, which commissioned the Klaipeda terminal at the end of the 2014 year. However, four new countries –Jordan, Egypt, Pakistan and Poland– are expected to join the LNG market in 2015, bringing the number of international importers to 33.

Among importers, the undisputed leader is Japan (Table 2) importing up to 37.6% of the total imported gas. Second place is South Korea, where the market share of imported gas is 16.1%,

Table 3. LNG trade between Basins in 2014 in MT (World LNG Report 2015).

Exporting Region	Africa	Asia Pacific	Europe	Former Soviet Union	Latin America	Middle East	North America	Re-exports Received	Re-exports Loaded	Total
Asia	3.2	9.7	0.2	0.1	0.2	20.5	-	0.6	-	34.5
Asia Pacific	13.0	63.8	0.4	10.5	0.9	54.4	0.3	2.4	(0.2)	145.5
Europe	15.3	-	1.9	-	3.5	17.7	-	0.6	(6.0)	33.0
Latin America	2.7	-	0.9	-	8.4	1.2	-	2.1	(0.1)	15.2
Middle East	0.5	0.3	-	-	0.9	2.2	-	0.4	-	4.3
North America	1.8	0.2	0.2	-	4.9	1.2	-	0.2	(0.1)	8.4
Total	36.6	74.0	3.6	10.6	18.8	97.2	0.3	6.4	(6.4)	241.1

Table 4. Global Liquefaction Plants in years 1969–2014 (own elaboration based on: World LNG Report 2015).

Country	Number of plants	Nameplate capacity in MPTA
Qatar	14	77,0
Indonesia	10	34,05
Australia	8	28,5
Nigeria	6	21,9
Algeria	6	27,2
Trinidad	4	15,5
Malaysia	4	23,9
Egypt	3	12,2
Oman	3	10,65
Russia	2	9,6
UAE	2	5,8
PNG	2	7,0
Yemen	2	6,7
Libya	1	3,2
Norway	1	4,2
Angola	1	5,2
Eq. Guinea	1	3,7
Peru	1	4,45
USA	1	1,5
Brunei	1	7,2
TOTAL	73	309,45

followed by China, India and Taiwan with a share of 8.5%, 6.2% and 5.8%.

The global LNG market can be simply divided into two areas: the Atlantic Basin and Pacific Basin. Currently, the main directions of LNG trade (Table 3) is export from the areas of Middle East (Qatar, Oman, Yemen, UEA) and from the areas of Pacific (Malaysia, Australia, Indonesia, Trinidad) to Asia, which accounts for 62% of global demand (China, Japan and South Korea). 40% of global demand for LNG covers the Middle East, while 31% of global demand covers the Pacific region.

At the end of 2014 in the world there were 73 LNG export terminals located in 20 countries with a total capacity of 309.45 MPTA (Table 4). In the next 5 years the construction will start of 26 consecutive (Table 5), most of which will be in

Australia (12) and the USA (6). In 101 terminals located in 30 countries (Table 6), the reception of supplied LNG has place. Over the next four years 25 consecutive terminals will be opened, including terminal in Poland in 2015 in Świnoujście (Table 7).

Table 5. Liquefaction plants under construction in years 2015–2019 (own elaboration based on: World LNG Report 2015).

Country	Number of projects	Nameplate capacity in MPTA
Australia	12	57,7
USA	6	44,05
Russia	3	16,5
Malaysia	3	6,3
Indonesia	1	2,0
Colombia	1	0,5
TOTAL	26	127,05

Table 6. LNG receiving terminals in years 1969–2014 (own elaboration based on: World LNG Report 2015).

Country	Number of terminals	Nameplate capacity in MPTA
Japan	23	188,2
China	12	39,4
USA	11	131,8
Spain	6	43,0
South Korea	5	98,1
UK	4	38,0
India	4	22,0
France	3	17,3
Mexico	3	16,7
Brazil	3	11,7
Italy	3	11,0
Taiwan	2	13,0
Turkey	2	10,3
Argentina	2	7,6
Indonesia	2	5,6
Chile	2	4,2
Netherlands	1	8,8
Canada	1	7,5
Belgium	1	6,6
Singapore	1	6,0
Portugal	1	5,8
Kuwait	1	5,8
Thailand	1	5,0
Malaysia	1	3,8
Greece	1	3,3
UAE	1	3,0
Israel	1	3,0
Lithuania	1	3,0
Dominican Rep.	1	1,9
Puerto Rico	1	1,2

Table 7. LNG receiving terminals under construction in years 2015–2018 (own elaboration based on: World LNG Report 2015).

Country	Number of projects	Nameplate capacity in MPTA
China	9	25,2
India	3	13,6
Japan	3	3,5
France	1	10,0
Egypt	1	3,8
Jordan	1	3,8
Poland	1	3,6
Indonesia	1	3,0
Uruguay	1	2,7
Pakistan	1	2,3
South Korea	1	2,0
Greece	1	1,9
Chile	1	1,3
TOTAL	24	73,7

4 MODELS OF LOGISTICS CHAINS FOR LIQUEFIED NATURAL GAS (LNG)

The supply chain for LNG can be described as a network of interrelated technical infrastructure and technological operations (Pielka 2013). The LNG infrastructure, from the point of gas extraction until its delivery to the final customer, consists of the following links (Kubiak 2010):

- Subsystem of gas production;
- Subsystem of processing raw gas suitable for LNG production, LNG produced in the process of liquefaction storage, its transportation to the shipping terminal and loading the ship;
- Subsystem of maritime transport;
- Subsystem of regasification and turn sea-supplied gas into the national transmission system.

Gas liquefaction plants and regasification terminals can be divided in terms of size and amount of transported gas into:

- high volumes—(more than 300 tons of LNG / day) for the delivery of international and intercontinental maritime,
- medium—(up to 300 tons of LNG / day) of regional importance, linked most often with the pipeline LNG distribution network or land transport (rail or road),
- small—(up to 20 tons of LNG / day) of local importance, related to the distribution network of land transport (mainly automotive).

Treating the above division as the base, it is possible to define three models logistics supply chains for LNG distribution—high-tonnage (I), medium-tonnage (II) and low-tonnage (III) (Pielka 2013).

I. High-tonnage model, macro logistics is related to the processes occurring in the global and national economy. It consists of the following logistics chains elements:

1. gas field,
2. condensation of gas—gas liquefaction plant,
3. LNG marine terminal—loading,
4. transport by sea—tanker,
5. LNG marine terminal—unloading,
6. regasification—regasification plant,
7. storage and distribution of gas by pipeline.

II. Medium-tonnage model, mezo logistics is associated with processes within the industrial sector of economy. It consists of the following logistics chains elements:

1. gas pipeline distribution,
2. gas liquefaction plant
3. land transport—rail,
4. installation of regasification
5. pipeline distribution.

III. Low-tonnage model, micro logistics is related to the processes occurring in the enterprise. It consists of the following logistics chains elements:

1. storage tanks,
2. land transport—road,
3. regasification plant,
4. pipeline distribution.

LNG supply chains are the global chains, which means that companies belonging to the chain are located throughout the world. The economic situation in different world regions: economic crises, currency fluctuations, prices of crude oil and natural gas, etc.; and the world political situation: political relations between the countries, the activities of international organizations and international communities, armed conflicts, etc.; influence directly on the various links of supply chain and, indirectly, on the functioning of entire logistics chain.

5 LIQUEFIED NATURAL GAS (LNG) TRANSPORTATION

The basic system for the gas transportation is the national, interstate or intercontinental pipeline network, which can be classified due to its maximum working pressure and due to the materials from which they were constructed.

But the gas pipeline system is not able to meet the global demand for natural gas. Countries such as Japan and South Korea, which now import natural gas, have been isolated from sources of raw materials supply due to the distance between countries—producers of natural gas or terrain prevented the

construction of gas pipeline infrastructure. The solution to this problem is the development of natural gas transport system in liquefied form—LNG.

Analyses carried out by the Institute of Gas Technology indicates that the maritime transport of liquefied natural gas has a higher profitability compared to the underwater pipeline at a distance of 700 nautical miles and in comparison with onshore pipeline at a distance of 2200 nautical miles (Motowidlak 2014).

LNG transport over long distances usually is carried out by the sea in specially constructed for this purpose ships called tankers. Modern fleet of LNG tankers at the end of 2014 consisted of 421 ships. Among operated fleet of almost 69% were membranships (289 units). Among the total number of gas tankers dominated ships were with a capacity from 90 000 to 170 000 m³–320 units. These smallest ones, up to 24 000 m³ was 24, and those whose capacity exceeds 170 000 m³–66. Usually, these are modern ships whose age does not exceed 10 years (269 units). Only 44 of them (10%) is in use for over 25 years (The LNG Industry 2014). The orders for new ships to carry liquefied gas in 2016 amounts to about 50 new ships, and in 2017 more than 30 tankers. In total in docks are built 144 units with tanks capacity of 2.2 million m³.

In 2014 it was carried out 4023 sailings, which is an increase with 25, compared to 2013 (3998 cruises). In all, about 4 023 loaded vessels were delivered in 2014 (The LNG Industry 2014):

- 1,524 – to Japan (1 532 in 2013),
- 559 – to South Korea (616 in 2013),
- 273 – to China (260 in 2013),
- 219 – to Taiwan (204 in 2013),
- 210 – to India (195 in 2013),
- 660 – to Europe (661 in 2013),
- 241 – to Argentina, Brazil and Chile (224 in 2013)
- 175 – to North America (171 in 2013),
- 101 – to Indonesia, Malaysia, Singapore and Thailand (82 in 2013),
- 61 – to Israel, Kuwait and Dubai (53 in 2013).

For the LNG preparation to international shipping is applied usually the special technology, according to which the natural gas extracted from deposits on land or on a shelf is supplied by pipeline to terminals located on the coast. The gas in these terminals is purified and liquefied. LNG from tanks is pumped to gas carriers by thermally insulated pipeline. Loading of 150 000 m³ of gas takes 10 to 12 hours. One international contract for the LNG supply is usually supported by 1–7 LNG carriers with a capacity of 20 to 140 thousands m³, depending on the capacity and length of the delivery route (Filin & Zakrzewski 2006).

6 PERSPECTIVES OF TRADE DEVELOPMENT AND USAGE OF LIQUEFIED NATURAL GAS (LNG)

Thanks to its properties, the LNG is obvious (though not the only one) alternative to other commonly used fuels. One of the applications is the use of LNG to power ships in port. Ships in ports produce large amounts of exhaust gases causing crossing their binding emission limits. One solution is to connect them to the electrical power supply from the mainland, the other, switching to LNG fuel. In the latter case, the local NO_x and SO_x emissions are reduced to the levels below those achieved by the first solution (Bagniewski 2016).

Another example of application is the use of LNG to power passenger and transportation ships. Engines powered by LNG reduce CO₂ emissions by 25–30 percent, compared to the emissions of diesel fuels. It is not surprising that LNG has become increasingly popular among ship-owners as a marine fuel. In the Nordic countries, especially Norway, the owners of ferries more often order the units with gas-power or dual. For some time now ships with dual power system (dual fuel) procure the largest shipping companies, operating the passenger and cargo vessels (Na morzu 2016).

LNG is also used in power plants. Existing gas-fired power plants provide 20% of world energy production. Due to increased availability of gas and more onerous environmental products of coal combustion is expected to significantly increase this share.

It is worth to mention that the LNG as a fuel is used in thousands of trucks with dual fuel engines. Mainly in the USA and Australia. Passenger cars use of CNG due to the too big and heavy tanks and problems with regasification process.

Properties of LNG are also used in maritime transport beyond using it as fuel to: boost air cooling the main engines, engine cooling through heat exchangers, cooling water systems, air-conditioning and refrigeration catering and cargo (Bagniewski 2016).

7 CONCLUSIONS

In conclusion it should be noted the following challenges in the near future in front of the global LNG market having a huge impact on the development and changes occurring on it (Smith 2016):

Asia is the key market for LNG. Japan accounts for approximately 35% of the global LNG market. Combined with South Korea, it accounts for ~50% of it, while Asia as a whole accounts for 75%. The gradual return of nuclear reactors in Japan will mean depleted demand for LNG, which has

accounted for ~50% of its generation mix in the last few years.

Even though Asian LNG demand accounts for a huge share of the global market, LNG comprises a very small piece of total Asian energy consumption. As countries such as China and India shift away from coal, LNG will be the go-to fuel to help fill this supply gap, spurring on a serious bout of demand.

While the extent of returning nuclear energy in Japan is still relatively unknown, should we see the return of 5–7 nuclear reactors by 2017, this would mean ~11.5 gigawatts of electricity returning to the grid. This could displace approximately 10.9 million tons of LNG the equivalent of 12% of the country's LNG imports last year.

Australia has invested the most into LNG, however, with an incredibly unfortunate sense of timing. It has invested around \$200 billion in the past few years to massively expand its LNG capacity through eight new projects.

Although supply is charging higher, demand is expected to do the same. Asia is expected to lead the charge, with emphasis on China and India. Oxford Energy's high case scenario projects that demand growth from the region could nearly double by 2030.

Russia is set to be the wildcard in relation to LNG demand in Europe. An increasing focus on renewables for power generation, in combination with higher natural gas costs, have meant that total natural gas demand in Europe is in structural decline. That said, European domestic production is concurrently in structural decline also.

Given that Australia and the US are looking to challenge the supremacy of Qatar as the global leader of LNG exports, it is looking to raise its

game. Qatar is adapting its contracts to become more competitive with its leading consumers—who are, not surprisingly: Japan, South Korea and India.

REFERENCES

- Bagniewski M., 2016. LNG – własności i zastosowanie, http://www.dnv.pl/Binaries/LNG%20wlasnosci%20i%20zastosowanie_tcm144520590.pdf.
- BP Energy Outlook 2016 edition, <https://www.bp.com/content/dam/bp/pdf/energy-economics/energy-outlook-2016/bp-energy-outlook-2016.pdf>.
- Filin S. & Zakrzewski B, 2006. Światowy handel skroplonym gazem ziemnym (LNG)—stan obecny i kierunki rozwoju, *Energetyka* 11 (629): 846–854.
- Kubiak K., 2010. Światowy rynek skroplonego gazu ziemnego, *Zeszyty Naukowe Uniwersytetu Przyrodniczo-Humanistycznego w Siedlcach*, Nr 87, Seria: Administracja i Zarządzanie: 99–112.
- Motowidlak U., 2014. Transport LNG drogą morską jako jeden z elementów łańcucha dostaw gazu, *Logistyka* 6: 950–954.
- Na morzu rośnie popularność napędu LNG, 2016, <http://logistyka.wnp.pl>.
- Pielka D., 2013. Model łańcucha logistycznego dystrybucji LNG, *LogiTrans—VII Konferencja Naukowo-Techniczna: Logistyka, Systemy transportowe, Bezpieczeństwo w transporcie*: 301–305.
- Smith M., 2016. 10 Challenges Faced by the Global LNG Market, <http://clipperdata.com/10-challenges-faced-by-the-global-lng-market/>.
- The LNG industry in 2014, International Group of Liquefied Natural Gas Importers.
- World LNG Report—2015 Edition, International Gas Union, http://www.igu.org/sites/default/files/node-page-field_file/IGU-World%20LNG%20Report-2015%20Edition.pdf.
- www.statista.com.

An IoT course for a computer science graduate program

Xing Liu & Orlando Baiocchi

Institute of Technology, University of Washington Tacoma, Tacoma, WA, USA

ABSTRACT: Internet of Things (IoT) has attracted tremendous interest from industry. Numerous major technological companies in the world now have IoT products. The academia has been supplying graduates for the IoT workforce via programs such as computer engineering, computer science, and electrical engineering. However, only a limited number of universities have dedicated courses for IoT and these courses focus on different aspects of IoT. This paper introduces an IoT course developed in the Institute of Technology, University of Washington Tacoma. The course is a graduate-level course targeting students in the cyber-physical track of the master of science in computer science and system program. The course will teach essential IoT concepts such as physical and logical design, architecture and functional blocks, communications protocols, enabling technologies, smart objects, system management, security, data analytics, and ethical and environmental impact, and development tools. The course adopts a very pragmatic approach with 40% of the time planned for learning theory, 60% of the time for learning popular industrial IoT hardware and software development tools. It is expected that students who have completed the course will have sufficient skills for industrial IoT application development.

1 INTRODUCTION

1.1 *IoT in industry*

The phrase “Internet of Things” (IoT) has become one of the hottest buzzwords in the technological world nowadays. Not only has IoT become a new research paradigm, but also an important line of products in major technological companies. For example, IEEE has started the IEEE Internet of Things (IoT) Initiative [1] which, among other work, is organizing the leading conference IEEE World Forum on Internet of Things [2]. In the meantime, IEEE launched a publication titled “IEEE Internet of Things Journal” in 2014 [3]. Microsoft’s response to the “IoT movement” is the Azure IoT Suite [4]—a cloud-based software system with preconfigured solutions which address common IoT scenarios. Amazon’s answer for IoT is AWS IoT [5]—a cloud platform which lets connected devices interact with each other and with cloud applications. Google [6], Oracle [7], IBM [8] and Salesforce [9] are also providing tools for IoT data collection and analysis. On the hardware front, Texas Instrument [10] and Intel [11] both provide building blocks and enabling technologies for IoT. CISCO [12] is actively involved in IoT development as well.

1.2 *IoT in academia*

In response to industrial needs for trained graduates, academic institutions have started offering

courses dedicated to IoT. For instance, Harvard University’s CS 144/244 teaches “Secure and Intelligent Internet of Things” [13]. The Harvard course uses a large number of devices such as depth sensing (Kinect and depth cameras) and Samsung Galaxy and watch in the course. North Carolina State University [14] has the course CSC 591 scheduled for Spring 2016. The name of the course is “Internet of Things: Applications and Implementation”. Columbia University [15] and California Polytechnic State University [16] both have a course titled “Internet of Things”. Columbia University’s IoT course number is E6765. California Polytechnic State University’s course is CSC 520. Apart from having IoT courses in standard degree curriculum, universities are offering training either online or via continuing education. For instance, University of California San Diego and Qualcomm are offering a six-course IoT certificate via Coursera [17]. University of Washington Seattle Professional and Continuing Education is developing a certificate on IoT [18] which will start in Autumn 2016.

The Institute of Technology in University of Washington Tacoma (UWT) has developed its own IoT course (TCSS 573) in Autumn 2015. The course has received Institute approval and is scheduled to be offered in Autumn 2016. The course is at the graduate-level and it targets students enrolled in the cyber-physical option of the master of science in computer science and systems degree program. This paper describes the details of the course (with some adjustment made to the tentative weekly schedule approved).

2 IOT COURSE BY THE UNIVERSITY OF WASHINGTON TACOMA

2.1 Course design

The UWT IoT course TCSS 573 is designed with certain objectives in mind. The main objective is that, after completing the course, the students will be able to design and develop IoT hardware and software, and utilize cloud tools to collect and analyze data for IoT applications.

The course is designed to have adequate breadth and depth in IoT theory. However, great emphasis is placed on practical skills. Students will design and build hardware, write code for embedded systems, customize and develop IoT cloud services using commercial products. It is envisioned that 40% of the class time will be spent on theory, but with 60% of the time for learning practical hardware and software development skills.

The course will be worth 5 credits and will be taught in 10 weeks because of the quarter system adopted by the University. The class will meet 4 hours a week with each meeting being 2 hours. The format of the class will be a mixture of lectures, study of commercial development tools, and class discussions. Each week will cover a special theoretical topic in IoT, plus practical examination of one of the key features of the IoT device development tools and software development tools. The course will have a project so that students can apply everything they learn in the course in one place and gain practical skills for future employment in industry. A textbook will be recommended [19]. However, large amount of course material will come from various additional references including the Web. Every student will be required to purchase an IoT device development kit from Texas Instrument [20] with the cost of around US\$30.

2.2 Course content

In order to give students the overall picture of IoT application development, the course begins with an introduction to the architectures of IoT applications described in Amazon's AWS IoT [21] and Microsoft's Azure IoT [22]. The purpose is to illustrate the physical components of an IoT application: the "things" or the devices, the IoT services, other cloud services, the applications, and how they relate to each other. Therefore, the students will see what they need to understand, what skills they need to learn in order to be able to develop IoT applications.

Then the course introduces the formal models for IoT architecture, the functional blocks, the communication models [19, 23]. These models represent an IoT system from different perspectives

in a more abstract way and facilitate the students' understanding of the principles and operations of an IoT system.

The CISCO IoT Reference Model [24] is introduced in the course next. CISCO conceptualized an IoT application into seven levels and defined "edge" and "edge computing" in the reference model.

The course then studies two of the IoT application domains: smart cities and smart environments [25, 26] so that students can see how IoT technologies can actually be used in the real world.

IoT requires numerous enabling technologies such as wireless sensor networks [27, 28]. The course examines sensors and sensor networks, and wireless technologies such as Wi-Fi, Bluetooth, and ZigBee.

Cloud computing [29], security [30], and big data analytics [31] are essential enabling technologies for IoT. The course covers these technologies together with the AWS IoT platform and Azure IoT Suite.

IoT communication protocols MQTT (Message Queuing Telemetry Transport) and CoAP (constrained application protocol) [32, 33] are used to address the special characteristics of communications between IoT connected devices. These protocols are discussed in the course. Embedded system concepts are also discussed [34].

IoT, as a network of connected devices, needs tools that can install, manipulate and delete devices, as well as tools that model network and device configuration and state data. This is called IoT system management [35]. The course dedicates some time to look at this aspect of IoT.

The course then moves on to discuss smart objects [36]. IoT can be seen as a loosely coupled,

Table 1. Weekly class plan.

Week	TCSS 573 Topics	Hands on
1	Physical & logical design	Project info
2	IoT architecture and models	IoT hardware
3	Application domains	IoT hardware
4	Wireless sensor networks	IoT hardware
5	Cloud services & security	IoT hardware
6	Data analytics	IoT cloud
7	IoT communication protocols	IoT cloud
8	System management	IoT cloud
9	Smart objects	IoT cloud
10	Ethical and environmental impact	IoT cloud

decentralized system of cooperating smart objects. Smart objects are “things” that can understand and react to their environment. Smart objects are ready to be connected and exchange information over the Web. This part of the course has an in depth look of smart objects.

Finally, the ethical and environmental impact of IoT will be discussed [37] in order to increase students’ awareness of the potential complications of adopting the IoT technology.

Table 1 is a weekly breakdown of the topics to be taught in the UWV’s TCSS 573.

3 IoT DEVELOPMENT TOOLS

3.1 Development tool selection

The criteria for selecting the tools are two folds: a) the tools should expose the students to the full spectrum of IoT application development, ranging from the “things” to the “cloud”; b) The tools should be commercial products. Based on the above criteria, Texas Instrument’s IoT kit is selected for device (the “thing” side) development. Amazon’s AWS IoT and Microsoft’s Azure IoT Suite are selected for development at the cloud side.

3.2 Texas instrument IoT kit

Texas Instrument’s IoT kit provides a internet-on-a-chip solution. The main component of the kit is an integrated circuit chip named CC3200 that integrates an ARM Cortex-M4 microcontroller and a WiFi network module, as shown in Figure 1. An IoT device can be developed with a single CC3200 chip which consumes very little power (two AA batteries can power the chip for over a year). The chip includes numerous serial and parallel interfaces and a 4-channel ADC. The Wi-Fi network module on CC3200 implements WiFi Internet-on-a-chip which includes an 802.11 b/g/n radio, baseband and MAC, embedded TCP/IP and TLS/SSL stacks, HTTP server, and multiple Internet protocols. The CC3200 chip can work as in station mode or in access point mode. Developers can attach external devices such as sensors to CC3200 and use the CC3200 SDK and the Code Composer Studio IDE to program the chip using the C language.

Students will learn to use the Kit to develop the hardware of the “things” or “smart objects”. They will develop the hardware of “things” that are connectable and ready to communicate over the Internet and exchange information over the Web. They will attached sensors to the microcontroller on the CC3200, develop embedded C code to implement various functions required by IoT so that the “things” can transmits data wirelessly to

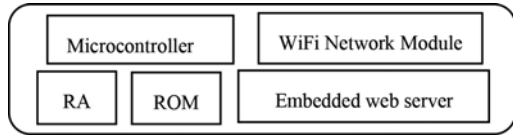


Figure 1. TI CC3200 Internet-on-a-chip SoC device.

the Internet and collaborate with other things and cloud services and applications.

3.3 Amazon AWS IoT

AWS IoT is the “middleman” between the “things” connected to the Internet and the AWS cloud. Users can collect data from the “things”, and store and analyze data in the cloud with the help of AWS IoT. Devices can interact with cloud applications and other devices. Users can also control the “things” via their mobile devices with the help of AWS IoT. With AWS IoT, developers can make use of standard AWS services such as: AWS Lambda which is a compute service that can receive user code and run the code on behalf of the user; Amazon Kinesis—a service that collects and processes data records in real time; Amazon S3—a simple web service interface for storing and retrieving data; Amazon Machine Learning which provides data visualization tools, machine learning models and predictions; as well as Amazon DynamoDB database service. AWS IoT provides different AWS SDKs for different hardware devices [37].

The learning experience for this tool will enable students to develop AWS IoT cloud services to communicate with the “things” they have developed using the CC3200 and visualize data and perform data analysis.

3.4 Microsoft Azure IoT Suite

Microsoft Azure IoT Suite is a cloud-based platform. It provides preconfigured solutions for common Internet of Things scenarios. Similar to AWS IoT, developers can use the Azure IoT Suite to connect “things”, to capture, manage, analyze and present data, and to automate operations. The center piece of the Azure IoT Suite is the Azure IoT Hub service. This service provides device-to-cloud and cloud-to-device messaging capabilities. It also acts as the gateway to the cloud and to the other key IoT Suite services.

Developers can use Azure Stream Analytics to perform data analysis and detect events and device responses to commands. Data can be stored using Azure Storage. Document DB can be used to manage device metadata. Data visualization can be achieved with Azure Web Apps and Microsoft

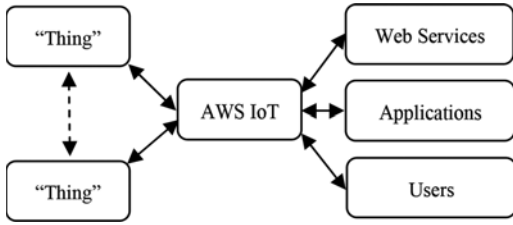


Figure 2. AWS system structure.



Figure 3. Microsoft Azure IoT.

Power BI. The preconfigured solutions and preconfigured services make IoT application development straightforward.

The learning experience for developing applications using the Azure IoT Suite [38] will equip the students with skills for developing IoT cloud services to communicate with the “things” they developed using the CC3200 and visualize data and perform data analysis, similar to the AWS IoT case.

4 CONCLUSIONS

This paper has introduced in detail the IoT graduate course developed in University of Washington Tacoma. The course is worth five-credits and lasts for one-quarter. It is designed to provide in-depth training in both IoT theory and IoT application development skills. Commercial hardware and software tools will be used in the course to equip students with practical skills much needed by the industry.

ACKNOWLEDGEMENT

The authors would like to thank Prof. Rajendra Katti for his encouragement in developing the course. Support from all members of the Graduate Curriculum Committee, Institute of Technology, University of Washington Tacoma is also greatly appreciated.

REFERENCES

[1] Iot.ieee.org, ‘About the IEEE Internet of Things (IoT) Initiative’, 2015. [Online]. Available: <http://iot.ieee.org/about.html>. [Accessed: 24-Dec-2014].

[2] Ieee-wf-iot.org, ‘IEEE 2nd World Forum on Internet of Things (WF-IoT)’, 2015. [Online]. Available: <http://www.ieee-wf-iot.org/>. [Accessed: 24-Dec-2014].

[3] Ieee.org, 2015. ‘Publications: IEEE Internet of Things Journal’. [Online]. Available: <http://iot-journal.weebly.com/>. [Accessed: 27-Dec-2014].

[4] Microsoft.com, ‘Tap into the Internet of Things with the Azure IoT Suite’, 2015. [Online]. Available: <http://www.microsoft.com/en-ca/server-cloud/internet-of-things/azure-iot-suite.aspx>. [Accessed: 24-Dec-2014].

[5] Aws.amazon.com, ‘AWS IoT’, 2015. [Online]. Available: <https://aws.amazon.com/iot/>. [Accessed: 24-Dec-2014].

[6] Cloud.google.com, ‘Internet of Things’, 2015. [Online]. Available: <https://cloud.google.com/solutions/iot/>. [Accessed: 24-Dec-2014].

[7] Oracle.com, ‘Oracle Internet of Things’, 2015. [Online]. Available: <https://www.oracle.com/solutions/internet-of-things/index.html>. [Accessed: 24-Dec-2014].

[8] Ibm.com, ‘Watson Internet of Things: IoT in the cognitive era’, 2015. [Online]. Available: <http://www.ibm.com/internet-of-things/watson-iot.html>. [Accessed: 24-Dec-2014].

[9] Salesforce.com, ‘Salesforce IoT Cloud’, 2015. [Online]. Available: <http://www.salesforce.com/ca/iot-cloud/>. [Accessed: 24-Dec-2014].

[10] Ti.com, ‘TI Internet of Things Overview’, 2015. [Online]. Available: http://www.ti.com/ww/en/internet_of_things/iot-overview.html. [Accessed: 24-Dec-2015].

[11] Intel.com, ‘The Internet of Things (IoT) Starts with Intel Inside’, 2015. [Online]. Available: <http://www.intel.com/content/www/us/en/internet-of-things/overview.html>. [Accessed: 24-Dec-2015].

[12] Cisco.com, ‘The Internet of Things’, 2015. [Online]. Available: <http://newsroom.cisco.com/internetofthings>. [Accessed: 24-Dec-2015].

[13] Harvard.edu, ‘CS 144r/244r: Secure and Intelligent Internet of Things’, 2014. [Online]. Available: <https://www.eecs.harvard.edu/htk/courses/>. [Accessed: 24-Dec-2015].

[14] Ncsu.edu, ‘CSC 591: Internet of Things: Applications and Implementation’, 2016. [Online]. Available: <http://www.csc.ncsu.edu/courses/>. [Accessed: 24-Dec-2015].

[15] Columbia.edu, ‘Course EECS E6765: Internet of Things’, 2014. [Online]. Available: <http://www.ee.columbia.edu/spring-2016-course-list>. [Accessed: 24-Dec-2015].

[16] Calpoly.edu, ‘CSC 520: Internet of Things’, 2015. [Online]. Available: <http://users.csc.calpoly.edu/~foaad/IOTS15.pdf>. [Accessed: 24-Dec-2015].

[17] Coursera.org, ‘Build your own Internet of Things’, 2015. [Online]. Available: <https://www.coursera.org/specializations/internet-of-things>. [Accessed: 24-Dec-2015].

[18] Uw.edu, ‘Certificate in Internet of Things’, 2015. [Online]. Available: <http://www.pce.uw.edu/certificates/internet-of-things.html>. [Accessed: 24-Dec-2015].

[19] Madiseti, V. & Bahga, A. 2014. Internet of Things (A Hands-on-Approach). VPT.

- [20] Ti.com, 'Ti's SimpleLink™ Wi-Fi® Family: Connect More: Anywhere, Anything, Anyone', 2015. [Online]. Available: http://www.ti.com/ww/en/simplelink_embedded_wi-fi/cc3200.html. [Accessed: 24-Dec-2015].
- [21] Amazon.com, 'How AWS IoT Works', 2015. [Online]. Available: <https://aws.amazon.com/iot/how-it-works/>. [Accessed: 24-Dec-2015].
- [22] S. George, 'Microsoft Azure IoT Suite—Connecting Your Things to the Cloud', 2015. [Online]. Available: <https://azure.microsoft.com/en-us/blog/microsoft-azure-iot-suite-connecting-your-things-to-the-cloud/>. [Accessed: 24-Dec-2015].
- [23] F. Carrez (editor), 'IoT Architecture: D1.5', 2013. [Online]. Available: www.ietf.org/public/public-documents/. [Accessed: 24-Dec-2015].
- [24] J. Green, 'Building the Internet of Things: CISCO internet of Things Reference Model', CISCO Connect, Oct. 2014, Manila, Philippines. DOI = https://www.cisco.com/web/PH/ciscoconnect/pdf/bigdata/jim_green_cisco_connect.pdf.
- [25] Cisco.com, 2013. "The Internet of Everything for Cities Connecting People, Process, Data, and Things To Improve the 'Livability' of Cities and Communities". [Online]. Available: <http://www.cisco.com/web/strategy/docs/gov/everything-for-cities.pdf>. [Accessed: 24-Dec-2015].
- [26] A. Zanella et al, 2014. Internet of Things for Smart Cities, *IEEE Internet of Things Journal*: 1(1), 22–32.
- [27] D. Culler et al. 2014. Overview of Sensor Networks, *IEEE Computer* 37(8): 41–49.
- [28] L. Mainetti, 2011. Evolution of wireless sensor networks towards the Internet of Things: A survey. *19th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, Adriatic Islands Split, Croatia, 1–6, Sept. 15–17.
- [29] A. Botta et al. 2014. On the Integration of Cloud Computing and Internet of Things, *International Conference on Future Internet of Things and Cloud (FiCloud)*: 23–30, Barcelona, Spain, Sept. 27–29, 2014.
- [30] R. Roman. et al, 2011. Securing the Internet of Things, *IEEE Computer* 44(9): 51–58.
- [31] R. Ciobanu. 2014. Big Data Platforms for the Internet of Things, in Bessis, Nik, Dobre, Ciprian (eds), *Big Data and Internet of Things: A Roadmap for Smart Environments*:3–34, Springer.
- [32] J. Stansberry, 2015. MQTT and CoAP: Underlying Protocols for the IoT. *Electronic Design*, Oct. 7, 2015. [Online]. Available: <http://electronicdesign.com/iot/mqtt-and-coap-underlying-protocols-iot>. [Accessed: 24-Dec-2015].
- [33] Eclipse.org, 2015. 'MQTT and CoAP, IoT Protocols'. [Online]. Available: https://eclipse.org/community/eclipse_newsletter/2014/february/article2.php. [Accessed: 26-Dec-2015].
- [34] W. Wolf, 2002. What is embedded computing? *IEEE Computer* 35(1): 136–137.
- [35] J. Schönwälder, 2012. Network Configuration Management with NETCONF and YANG, *84th IETF Meeting*, 29 July. Vancouver. [Online]. Available: <https://www.ietf.org/slides/slides-edu-netconf-yang-00.pdf>. [Accessed: 24-Dec-2015].
- [36] G. Kortuem et al. 2009. Smart objects as building blocks for the Internet of things, *IEEE Internet Computing* 14(1): 44–51.
- [37] R. Weber, 2010. Internet of Things—New security and privacy challenges, *Computer Law & Security Review* 26(1): 23–30.
- [38] Amazon.com, 2015. 'Getting Started'. [Online]. Available: <https://aws.amazon.com/iot/getting-started/#kits>. [Accessed: 24-Dec-2015].
- [39] Microsoft.com, 2015. 'OS Platforms and hardware compatibility with device SDKs'. [Online]. Available: <https://azure.microsoft.com/en-us/documentation/articles/iot-hub-tested-configurations/>. [Accessed: 24-Dec-2015].



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

An innovative knowledge discovery mechanism for unique pattern

Khaled ElBahnasy

Information Systems Department, Faculty of Computer and Information Sciences, Ain Shams University, Abbasia, Cairo, Egypt

ABSTRACT: Classification is considered one of the most important techniques used in Data Mining, A typical pattern recognition system consists of three phases. These are data acquisition, feature extraction and classification. This paper will discuss a new mechanism to get unique features using power set represented in vectors as a data structure. This mechanism faced limitation in the number of set items. The new technique is explained by algorithm and illustrated by example.

Keywords: Data Mining; Preprocessing; Classification; Power Set; Vector

1 INTRODUCTION

Knowledge discovery and data mining has recently emerged as an important research direction for extracting useful information from vast repositories of data of various types [1]. This chapter discusses some of the basic concepts and issues involved in this process with special emphasis on different data mining tasks. The major challenges in data mining are mentioned. Finally, the recent trends in data mining are described and an extensive bibliography is provided.

Data mining has attracted a great deal of attention in the information industry and in society as a whole in recent years [2], due to the wide availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge.

The information and knowledge gained can be used for applications ranging from market analysis, fraud detection, and customer retention, to production control and science exploration.

Data mining can be viewed as a result of the natural evolution of information technology. The database system industry has witnessed an evolutionary path in the development of the following functionalities (Figure 1): data collection and database creation, data management (including data storage and retrieval, and database transaction processing), and advanced data analysis (involving data warehousing and data mining). For instance, the early development of data collection and database creation mechanisms served as a prerequisite for later development of effective mechanisms for data storage and retrieval, and query and transaction processing. With numerous database systems offering query and transaction processing as com-

mon practice, advanced data analysis has naturally become the next target.

Essentially, the task of knowledge discovery can be classified into data preparation [1], data mining and knowledge presentation. Data mining is the core step where the algorithms for extracting the useful and interesting patterns are applied. In this sense, data preparation and knowledge

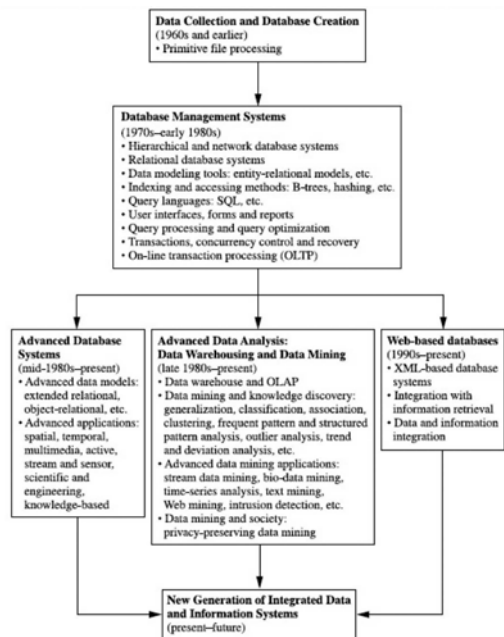


Figure 1. Steps in the process of knowledge discovery [26].

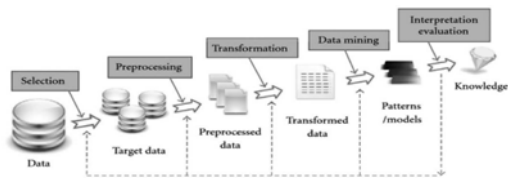


Figure 2. The knowledge discovery process [27].

presentation can be considered, respectively, to be pre-processing and post-processing steps of data mining. Figure 2 presents a schematic view of the steps involved in the process of knowledge discovery. The different issues pertaining to KDD are now described.

Section 2 of this paper describes the definition of data mining, All Data mining techniques and the difference between them. Section 3 proposes the highlight of the classification technique used in data mining, How to make an efficient classification. Then, section 4 presents what is power set and how it can be used to get the unique combinations using different approaches and difference between them. Finally, section 5 presents the Conclusion & Future Work.

2 DATA MINING

Data mining is formally defined as the process of discovering interesting, previously unknown and potentially useful patterns from large amounts of data. Patterns discovered could be of different types such as associations, sub-graphs, changes, anomalies and significant structures. It is to be noted that the terms interesting and potentially useful are relative to the problem and the concerned user. A piece of information may be of immense value to one user and absolutely useless to another. Often data mining and knowledge discovery are treated as synonymous, while there exists another school of thought which considers data mining to be an integral step in the process of knowledge discovery. Data mining techniques mostly consist of three components [3]; a model, a preference criterion and a search algorithm. The most common model functions in current data mining techniques include classification, clustering, regression, sequence and link analysis and dependency modelling. Model representation determines both the flexibility of the model for representing the underlying data and the interpretability of the model in human terms. This includes decision trees and rules, linear and nonlinear models, example-based techniques such as NN-rule and case-based reasoning, probabilistic graphical dependency models (e.g., Bayesian

network) and relational attribute models. The preference criterion is used to determine, depending on the underlying data set, which model to use for mining, by associating some measure of goodness with the model functions. It tries to avoid over-fitting of the underlying data or generating a model function with a large number of degrees of freedom. Finally, once the model and the preference criterion are selected, specification of the search algorithm is defined in terms of these along with the given data.

2.1 Data pre-processing

Data available for mining is raw data. Data may be in different formats as it comes from different sources, it may consist of noisy data, irrelevant attributes, missing data etc. Data needs to be pre-processed before applying any kind of data mining algorithm which is done using following steps [15]:

Data Integration—If the data to be mined comes from several different sources data needs to be integrated which involves removing inconsistencies in names of attributes or attribute value names between data sets of different sources.

Data Cleaning—This step may involve detecting and correcting errors in the data, filling in missing values, etc. Some data cleaning methods are discussed in [16, 17].

Discretization—When the data mining algorithm cannot cope with continuous attributes, discretization needs to be applied. This step consists of transforming a continuous attribute into a categorical attribute, taking only a few discrete values. Discretization often improves the comprehensibility of the discovered knowledge [18,19].

Attribute Selection—not all attributes are relevant so for selecting a subset of attributes relevant for mining, among all original attributes, attribute selection is required.

2.2 Data mining techniques

In this section, several data mining techniques and the difference between them are going to be shown, later on the classification techniques using unique features of each set of data will be concise in this paper.

2.2.1 Classification

Classification is considered one of the most important and crucial tasks in data mining, A typical pattern recognition system consists of three phases. These are data acquisition, feature extraction and classification. In the data acquisition phase, depending on the environment within which the

objects are to be classified, data are gathered using a set of sensors. These are then passed on to the feature extraction phase, where the dimensionality of the data is reduced by measuring/retaining only some characteristic features or properties. In a broader perspective, this stage significantly influences the entire recognition process. Finally, in the classification phase, the extracted features are passed on to the classifier that evaluates the incoming information and makes a final decision. This phase basically establishes a transformation between the features and the classes.

The problem of classification is basically one of partitioning the feature space into regions, one region for each category of input. Thus it attempts to assign every data point in the entire feature space to one of the possible (say, k) classes. Classifiers are usually, but not always, designed with labelled data, in which case these problems are sometimes referred to as supervised classification (where the parameters of a classifier function D are learned). Some common examples of the supervised pattern classification techniques are the Nearest Neighbour (NN) rule, Bayes maximum likelihood classifier and perceptron rule [4, 5, 6, 7, 8, 9, 10, 11, 12].

Figure 3 provides a block diagram showing the supervised classification process. Some of the related classification techniques are described below. NN Rule [6, 8, 12] Let us consider a set of n pattern points of known classification $\{x_1, x_2, \dots, x_n\}$, where it is assumed that each pattern belongs to one of the classes C_1, C_2, \dots, C_k . The NN classification rule then assigns a pattern x of unknown classification to the class of its nearest neighbour, where $x_i \in \{x_1, x_2, \dots, x_n\}$ is defined to be the nearest neighbour of x if

$$D(x_i, x) = \min \{D(x_j, x)\}, 1 = 1, 2, \dots, n$$

where D is any distance measure definable over the pattern space.

2.2.2 Association

Association (or relation) is probably the better known and most familiar and straightforward

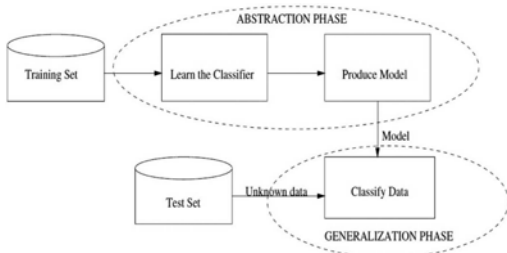


Figure 3. The supervised classification process.

data mining technique. Here, you make a simple correlation between two or more items, often of the same type to identify patterns. For example, when tracking people's buying habits, you might identify that a customer always buys cream when they buy strawberries, and therefore suggest that the next time that they buy strawberries they might also want to buy cream [13].

Building association or relation-based data mining tools can be achieved simply with different tools. For example, within InfoSphere Warehouse a wizard provides configurations of an information flow that is used in association by examining your database input source, decision basis, and output information.

2.2.3 Clustering

By examining one or more attributes or classes, you can group individual pieces of data together to form a structure opinion [13]. At a simple level, clustering is using one or more attributes as your basis for identifying a cluster of correlating results. Clustering is useful to identify different information because it correlates with other examples so you can see where the similarities and ranges agree.

Clustering can work both ways. You can assume that there is a cluster at a certain point and then use identification criteria to see if you are correct. The graph in Figure 4 shows a good example. In this example, a sample of sales data compares the age of the customer to the size of the sale. It is not unreasonable to expect that people in their twenties (before marriage and kids), fifties, and sixties (when the children have left home), have more disposable income.

In the example, two clusters can be identified, one around the US\$2,000/20–30 age group, and another at the US\$7,000–8,000/50–65 age group. In this case, the hypothesis has been both

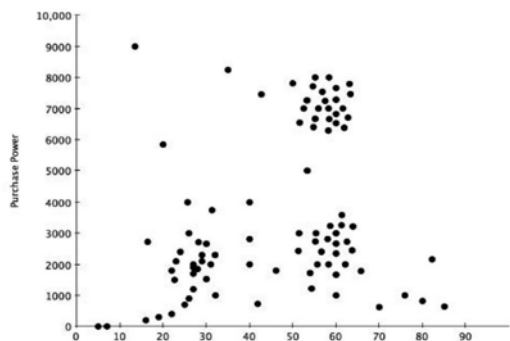


Figure 4. Clustering.

hypothesized and proved with a simple graph that creates using any suitable graphing software for a quick manual view. More complex determinations require a full analytical package, especially if you want to automatically base decisions on *nearest neighbor* information.

Plotting clustering in this way is a simplified example of so called *nearest neighbor* identity. You can identify individual customers by their literal proximity to each other on the graph. It's highly likely that customers in the same cluster also share other attributes and you can use that expectation to help drive, classify, and otherwise analyze other people from your data set.

You can also apply clustering from the opposite perspective; given certain input attributes, you can identify different artifacts. For example, a recent study of 4-digit PIN numbers found clusters between the digits in ranges 1–12 and 1–31 for the first and second pairs. By plotting these pairs, you can identify and determine clusters to relate to dates (birthdays, anniversaries).

2.2.4 Prediction

Prediction is a wide topic and runs from predicting the failure of components or machinery, to identifying fraud and even the prediction of company profits [13]. Used in combination with the other data mining techniques, prediction involves analyzing trends, classification, pattern matching, and relation. By analyzing past events or instances, you can make a prediction about an event.

Using the credit card authorization, for example, you might combine decision tree analysis of individual past transactions with classification and historical pattern matches to identify whether a transaction is fraudulent. Making a match between the purchase of flights to the US and transactions in the US, it is likely that the transaction is valid.

2.2.5 Sequential patterns

Often used over longer-term data, sequential patterns are a useful method for identifying trends, or regular occurrences of similar events [13]. For example, with customer data you can identify that customers buy a particular collection of products together at different times of the year. In a shopping basket application, you can use this information to automatically suggest that certain items be added to a basket based on their frequency and past purchasing history.

2.2.6 Combinations

In practice, it's very rare that you would use one of these exclusively. Classification and clustering are similar techniques. By using clustering to identify nearest neighbors, you can further refine your classifications [13]. Often, the decision trees are used

to build and identify classifications that can be tracked for a longer period to identify sequences and patterns.

2.2.7 Long-term (memory) processing

Within all of the core methods, there is often reason to record and learn from the information. In some techniques, it is entirely obvious [13]. For example, with sequential patterns and predictive learning you look back at data from multiple sources and instances of information to build a pattern.

In others, the process might be more explicit. Decision trees are rarely built one time and are never forgotten. As new information, events, and data points are identified, it might be necessary to build more branches, or even entirely new trees, to cope with the additional information.

You can automate some of this process. For example, building a predictive model for identifying credit card fraud is about building probabilities that you can use for the current transaction, and then updating that model with the new (approved) transaction. This information is then recorded so that the decision can be made quickly the next time.

3 CLASSIFICATION

The classification task can be seen as a supervised technique where each instance belongs to a class, which is indicated by the value of a special goal attribute or simply the class attribute [20].

The goal attribute can take on categorical values, each of them corresponding to a class.

3.1 Rule based classifiers

Rule based classifiers deals with the discovery of high-level, easy-to-interpret classification rules of the form if-then.

3.2 Bayesian Networks

A Bayesian Network (BN) consists of a directed, acyclic graph and a probability distribution for each node in that graph given its immediate predecessors [21]. A Bayes Network Classifier is based on a bayesian network which represents a joint probability distribution over a set of categorical attributes.

3.3 Decision trees

Related to most of the other techniques (primarily classification and prediction), the decision tree can be used either as a part of the selection criteria, or to support the use and selection of specific

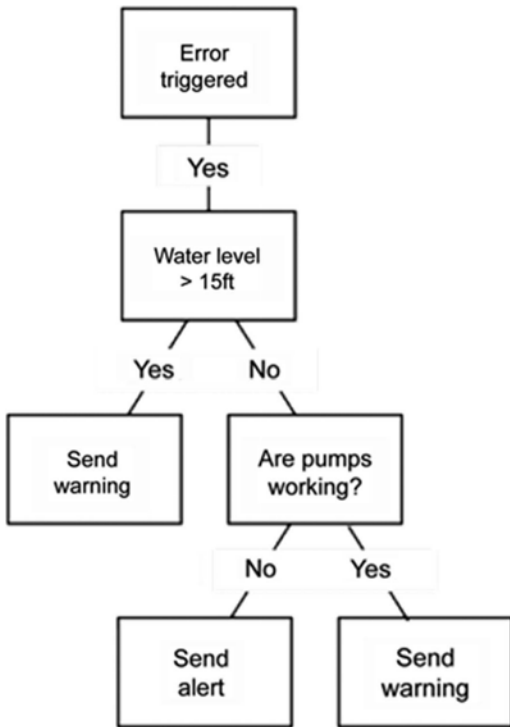


Figure 5. Decision tree.

data within the overall structure [13]. Within the decision tree, you start with a simple question that has two (or sometimes more) answers. Each answer leads to a further question to help classify or identify the data so that it can be categorized, or so that a prediction can be made based on each answer.

Figure 5 shows an example where you can classify an incoming error condition.

Decision trees are often used with classification systems to attribute type information, and with predictive systems, where different predictions might be based on past historical experience that helps drive the structure of the decision tree and the output.

3.4 Nearest Neighbor

A Nearest Neighbor Classifier assumes all instances correspond to points in the n -dimensional space. During learning, all instances are remembered. When a new point is classified, the k -nearest points to the new point are found and are used with a weight for determining the class value of the new point. For the sake of increasing accuracy, greater weights are given to closer points [22].

3.5 Artificial Neural Network

An artificial neural network, often just called a neural network is a mathematical model or computational model based on biological neural networks, in other words, is an emulation of biological neural system. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase [23].

3.6 Support Vector Machines

Support Vector Machines [24] are basically binary classification algorithms. Support Vector Machines (SVM) is a classification system derived from statistical learning theory. It has been applied successfully in fields such as text categorization, hand-written character recognition, image classification, biosequences analysis, etc. The SVM Separates the classes with a decision surface that maximizes the margin between the classes.

3.7 Rough sets

Any set of all indiscernible (similar) objects is called an elementary set. Any union of some elementary sets is referred to as a crisp or precise set—otherwise the set is rough (imprecise, vague). Each rough set has boundary-line cases, i.e., objects which cannot be with certainty classified, by employing the available knowledge, as members of the set or its complement [25].

3.8 Fuzzy logic

Fuzzy logic is a multivalued logic different from “crisp logic”, where binary sets have two valued logic. Fuzzy logic variables have truth value in the range between 0 and 1. Fuzzy logic is a superset of conventional Boolean logic that has been extended to handle the concept of partial truth.

3.9 Genetic Algorithms

Genetic Algorithms (GA) are search algorithms based on natural genetics that provide robust search capabilities in complex spaces, thereby offering a valid approach to problems requiring efficient and effective search processes.

4 UNIQUE FEATURES USING POWER SET

This section covers the history of the proposed work and the problems which appeared throughout the implementation of this technique.

Before explaining the algorithm steps, it is necessary to show some terminologies and symbols that will be used.

C = Case, **A** = Attributes, **CL** = Class Label, **S** = Set of Case Attributes, **OKB** = Output Knowledge Base, **DS** = Data Set.

Subsumsion = Set is entailed from another smaller set, i.e.: $S1 = \{A, B\}$, $S2 = \{A, B, C\}$, **S2** is Subsumsion from **S1**.

The following Algorithm-1 illustrates how get the unique feature for each case.

Algorithm-1:

1. Sort all cases ascending in DS according to number of attributes.
2. For Each Case " C_i " in the DS.
 - 2.1 Read Case " C_i ".
 - 2.2 Generate Power Set for its Attributes " A_i ".
 - 2.3 Loop for each Combination CM_i .
 - 2.3.1 If (Combination CM_i is Unique in DS).
 - 2.3.1.1 IsSubSumsion (CM_i, OKB)
 - 2.3.1.2 If Yes, Go to Step 2.3.
 - 2.3.1.3 If No, Add CM_i with its class label CL_i to the OKB.
3. Go to Step 2.

The vector data structure is used to represent power set of each case. The following sample shows how to apply this algorithm on DS to get the unique features using vector data structure.

Suppose the DS contains the following cases as shown in Table 1.

After Sorting the DS, the cases would be as the following Table 2.

Initially the output knowledge base (OKB) is empty, so the following tables illustrate the steps of the algorithm.

Case No. = C_2

Attributes Set = $[A_1, A_4]$

Power Set of Attributes Set = $[[A_1],[A_4],[A_1, A_4]]$

The subset $[A_1]$ is not unique in DS because it exists for CL_1 AND CL_5

Table 1. Data set example.

#	Attributes	Lass label
C_1	A_3, A_4, A_6, A_7	CL_1
C_2	A_1, A_4	CL_1
C_3	A_6, A_8, A_9	CL_2
C_4	A_5, A_6, A_7	CL_3
C_5	A_2, A_8	CL_3
C_6	A_2, A_3, A_5	CL_3
C_7	A_3, A_7, A_{10}	CL_4
C_8	A_1, A_7, A_{10}	CL_5

Table 2. Sorted data set.

#	Attributes	Lass label
C_2	A_1, A_4	CL_1
C_5	A_2, A_8	CL_3
C_3	A_6, A_8, A_9	CL_2
C_4	A_5, A_6, A_7	CL_3
C_6	A_2, A_3, A_5	CL_3
C_7	A_3, A_7, A_{10}	CL_4
C_8	A_1, A_7, A_{10}	CL_5
C_1	A_3, A_4, A_6, A_7	CL_1

The subset $[A_4]$ unique in DS AND does not exist in OKB, thus, the following new rule will be add to OKB

R1: $A_4 \longrightarrow CL_1$

The subset $[A_1, A_4]$ contains $[A_4]$ which is a unique Attribute.

Applying the previous steps for C_5 will produce the following new rule

R2: $A_2 \longrightarrow CL_2$

Applying the previous steps for C_3 will lead to the following new rules

R3: $A_9 \longrightarrow CL_2$

R4: $A_6, A_8 \longrightarrow CL_2$

Applying the previous steps for C_4 will lead to the following new rules

R5: $A_5 \longrightarrow CL_3$

Applying the previous steps for C_6 will lead to nothing.

Applying the previous steps for C_7 will lead to the following new rules

R6: $A_3, A_{10} \longrightarrow CL_4$

Applying the previous steps for C_8 will lead to the following new rules

R7: $A_1, A_7 \longrightarrow CL_5$

R8: $A_1, A_{10} \longrightarrow CL_5$

Applying the previous steps for C_1 will lead to the following new rules

Finally the Output knowledge base contains the following rules

R1: $A_4 \longrightarrow CL_1$

R2: $A_2 \longrightarrow CL_2$

R3: $A_9 \longrightarrow CL_2$

R4: $A_6, A_8 \longrightarrow CL_2$

R5: $A_5 \longrightarrow CL_3$

R6: $A_3, A_{10} \longrightarrow CL_4$

R7: $A_1, A_7 \longrightarrow CL_5$

R8: $A_1, A_{10} \longrightarrow CL_5$

R9: $A_3, A_6 \longrightarrow CL_1$

As shown in the previous sample, number of comparisons is equal to the number of combinations generated $(2^n - 1)$, Two checks have to be performed for each finding, one for uniqueness and the other is for Subsumsion.

From real case with different data sets, by using Core 2 Duo CPU and 4 GB Ram, it is impossible to work on more than one case with 20 finding ($2^{20} - 1$). As performing all these computations consumed all available memory and CPU.

5 CONCLUSION AND FUTURE WORK

In this paper data mining techniques is discussed and it highlighted the new classification mechanism used in data mining. Then it shows how to get unique features using power set by getting unique combinations can identify any case. The vector as a data structure is used for representing the power set of each case in the data set. Using the vector as a data structure leads to limitation in the mechanism causing an out of memory exception for any case has more than 20 attributes. In the future this mechanism needs a new data structure to solve this limitation and to be implemented on parallel processors that will affect the performance greatly and will help in getting the unique combinations for larger attributes.

REFERENCES

- [1] Fayyad, U., G. Piatetsky-Shapiro and P. Smyth, 1996: The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, **39**, 27–34.
- [2] Jiawei Han, Micheline Kamber, *Data Mining Concepts and Techniques, Second Edition*.
- [3] Andrews, H. C., 1972: *Mathematical Techniques in Pattern Recognition*. Wiley Interscience, New York.
- [4] Devijver, P. A. and J. Kittler, 1982: *Pattern Recognition: A Statistical Approach*. Prentice-Hall, London.
- [5] Duda, R. O. and P. E. Hart, 1973: *Pattern Classification and Scene Analysis*. John Wiley, New York.
- [6] Fu, K. S., 1982: *Syntactic Pattern Recognition and Applications*. Academic Press, London.
- [7] Fukunaga, K., 1972: *Introduction to Statistical Pattern Recognition*. Academic Press, New York.
- [8] Gelsema, E. S. and L. N. Kanal, eds., 1986: *Pattern Recognition in Practice II*. North Holland, Amsterdam.
- [9] Gonzalez, R. C. and M. G. Thomason, 1978: *Syntactic Pattern Recognition: An Introduction*. Addison-Wesley, Reading, MA.
- [10] Pavlidis, T., 1977: *Structural Pattern Recognition*. Springer-Verlag, New York.
- [11] Tou, J.T. and R.C. Gonzalez, 1974: *Pattern Recognition Principles*. Addison-Wesley, Reading, MA.
- [12] Anderson, T. W., 1958: *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.
- [13] *Fundamentals of contemporary set theory. Universitext. Springer-Verlag. ISBN 0-387-90441-7. Zbl 0407.04003.*
- [14] Z.Z. Shi, Knowledge discovery, Tsinghua University Press, Beijing, 2001.
- [15] D. Pyle, Data preparation for data mining, 1st Vol., Morgan Kaufmann publisher, San Francisco, 1999.
- [16] I. Guyon, N. Matic and V. Vapnik, “Discovering informative patterns and data cleaning”, In: Fayyad UM, Piatetsky-Shapiro G, Smyth P and Uthurusamy R. (ed) *Advances in knowledge discovery and data mining*, AAAI/MIT Press, California, 1996, pp. 181–203.
- [17] Piatetsky-Shapiro G, Smyth P and Uthurusamy R., “Integrating inductive and deductive reasoning for data mining”, In: Fayyad UM, *Advances in knowledge discovery and data mining*, AAAI/MIT Press, California, 1996, pp. 353–373.
- [18] B. Pfahringer, “Supervised and unsupervised discretization of continuous features”, *Proc. 12th Int. Conf. Machine Learning*, 1995, pp. 456–463.
- [19] J. Catlett, “On changing continuous attributes into ordered discrete attributes”, In Y. Kodratoff (ed), *Machine Learning—EWSL-91*, Springer-Verlag, New York, 1991, pp. 164–178.
- [20] Sunita Beniwal, Jitender Arora, *Classification and Feature Selection Techniques in Data Mining*, Department of Information Technology, Maharishi Markandeshwar University, Mullana, August 2012, Ambala-133203, India.
- [21] A. Darwiche, *Modeling and Reasoning with Bayesian Networks*, Cambridge University Press, 2009.
- [22] T.M. Mitchell, *Machine Learning*, McGraw-Hill Companies, USA, 1997.
- [23] Y. Singh Y, A.S. Chauhan, “Neural Networks in Data Mining”, *Journal of Theoretical and Applied Information Technology*, 2005, pp. 37–42.
- [24] V. N. Vapnik, *Statistical Learning Theory*, Wiley New York., 1998.
- [25] Yumin Chen, Duoqian Miao, Ruizhi Wang, Keshou Wu, A rough set approach to feature selection based on power set tree, *Knowledge-Based Systems*, 2011, 275–281.
- [26] Han, J. and Kamber, M., *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 2000/2006.
- [27] Ozlem Terzi, “Monthly Rainfall Estimation Using Data-Mining Process”, *Applied Computational Intelligence and Soft Computing*, 2012.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

The degree of knowledge management practice in the department of legal affairs: Saudi Arabian Airlines case study

Abdullah M.A. Al-Yateem & Nawaf B. bin Hamid

Information Science Department, King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT: The main objective of this article is to identify the degree of knowledge management practice in the department of legal affairs in Saudi Arabian Airlines, the researcher followed the descriptive approach “case study” by using a questionnaire as a tool to collect data; the study community was the department of legal affairs section in Saudi Arabian Airlines and the research sample consisted of 4 heads of the sections in this department. The importance of this study relies on that it tackles a recent topic which concerns the business core of the leading managements in Saudi Arabian Airlines as these management works in an environment characterized by evolving, changing and renewing. Also, the knowledge management is one of the basic means that enables this department of reaching to the excellent stage in achieving its objectives. The study came at that the degree of management practice in the field of gaining knowledge reached 55, 71% with average index, in the field of generating knowledge reached 38% with low index, in the field of spreading knowledge reached 40% with low index, in the field of exchanging and sharing knowledge reached 47, 69% and in the field of knowledge techniques reached 35,38% with low index. Generally, the degree of knowledge management practice in the legal department reached 43,36% with low index.

The study suggested that the department can attract the experienced to give lectures to employees and those who have an academic experience in the field of business administration to develop the employees’ performance. The department should do research and creativity competitions and motivate the excellent people with rewards. The study also confirms that the department should cooperate with human resources in adopting creative and innovative performance and with community institutions in adopting creative people. Also, the department brings together the mass media methods to play its role in supporting the creative. The financial bonuses and expenses related to generating knowledge should be included in the department budget. For spreading knowledge, the management has to utilize its website and create accounts on social media. Organizing lectures and seminars inside and outside the department would be also of a good benefit besides attracting knowledge experts to sustain the knowledge work in the legal department. Finally, it should be seriously thought in creating a new section to be called knowledge management which undertakes the operations of knowledge management and its concerning. The study recommended building a suggestion of the knowledge management in the legal department in Saudi Arabian Airlines and making the same kind of studies on the rest of Saudi Arabian Airlines departments and managements. Finally, it’s recommended making a study on knowledge management planning in general in Saudi Arabian Airlines.

Keywords: knowledge management, Saudi Arabian management, department of legal affairs, knowledge management practice, descriptive approach “case study”, SPSS

1 INTRODUCTION

The knowledge management gains an increasing importance under large challenges which faces the organizations and this importance increases under the increase of the importance of knowledge objectives which the knowledge management focuses on achieving to promote the productivity, efficiency and effectiveness levels in the organizations. To achieve the desired benefit from authorizing the

introduction to knowledge management in organizations, the role of the organization management should focus on the effective use of this introduction through employing it to achieve the organization strategic and operative objectives, promoting the various capacities of the organization and its cadres’ skills and achieving development, improvement and continuity of these skills and capacities. The organization management should focus on directing the operations of knowledge manage-

ment for execution and dedication of institutionalization (knowledge institutionalization). Also, it should focus on applying a knowledge strategy that ensures the effectiveness of knowledge management operations in all organization units integrally.

The study tackles this renewed active topic on a group of vital institutions in East Jerusalem (civil institutions) for the purpose of focusing on the importance of applying this contemporary administrative introduction on these institutions. Civil institutions in East Jerusalem means the non governmental institutions, clubs and associations that offers public and specialized services in social, healthy, educational, democratic and human rights fields. The excellence center defines the civil institutions (2002) as any non profit union, association, institution, charitable fund or foundation (corporation). The researcher would review the study problem, queries, solution methodology, literature, digital outcomes and its interpretations in addition to presenting the final outcomes, suggestions and recommendations at the end.

2 LITERATURE REVIEW

The first study [1] try to identify the relationship between the requirements of knowledge management and its operations and its impact on the excellence of institutional performance. Also, the study seeks to explore the relationship between the requirements of knowledge management (knowledge needs, awareness and commitment, inside and outside communications) and knowledge management operations (personification, generation, storage, and distribution, application) and the effect of this relationship on the excellence of the institutional performance at the Jordanian High Education Ministry. For achieving the study objectives, the researcher made a questionnaire to collect primary data which consists of 90 paragraphs and distribute it on the staff of study sample which is the employees with intermediate diploma and above (300 employee). Thus, the data has been collected and analyzed and hypothesis has been tested using SPSS. The study found that it is important to have a statistical relationship between knowledge management requirements and each of its operations (personification, generation, storage, distribution, application) on one side and to have a statistical relationship between the requirements of knowledge management and its operations and worker satisfaction, learning, institutional growing and efficiency of internal operations on the other side.

The study [2] aimed at identifying the degree of practicing knowledge management opera-

tions (organizing, generating, participating, and applying) at the faculty of Education in Taif. This study used the analytical descriptive approach; its community and sample would be all of the teaching staff besides using the cadastral descriptive approach. The study came at that the four operations include positive and negative practices and the descending order of relative importance to knowledge management operations is: organizing (0,67), generating (0,67), participating (0,63) and applying (0,56). The study gives many recommendations such as providing financial support necessary for supporting and activating knowledge operations, connecting the faculty with local and international scientific research centers for participating in knowledge and performing more studies.

The study [3] aimed at presenting an intellectual framework for applying knowledge management concept in high education institutions depending on reviewing and checking a number of theoretical studies and applicable experiments of other high education institutions over the world. The study includes 3 topics: the first one tackles basic concepts of knowledge management (concept and kinds of knowledge, knowledge management concept, and basic items of knowledge management, knowledge objectives, benefits and requirements of application). The second topic titled knowledge management in high education institutions tackles many items (definition of knowledge management in high education institutions, reasons of application, expected benefits after applying it in these institutions, fields and steps of application). The third topic reviewed 3 experiments of applied knowledge management in high education institutions: Yung Ta Institute of Technology and Commerce experiment in Taiwan, Plymouth university experiment in Britain and University of Malaya experiment in Malaysia. Thus, the most important factors helping in achieving success and challenges that may appear when applying knowledge management in high education institutions have been reviewed. The study concluded with a summary having the most important discussed topics besides some recommendation such as a suggestion on application steps of knowledge management in high education institutions and other recommendations which shows the researcher's point of view about applying knowledge management in high education institutions.

The study [4] aimed at defining the role of knowledge management in improving and promoting the performance level; the knowledge management highlights and utilizes the valuable information besides understanding knowledge assets of the organization and how to make use of them. Knowledge management is considered as a base for the sake of survival, excellence and perfec-

tion because of its impact on the performance. It has been applied on ten modern companies established under the Investment Encouragement Act 1991. Data have been collected basically through a questionnaire and interviews set for this purpose for five years (2003–2007). The study came at outcomes which show that there is a great connection between knowledge management and performance and this helps the researcher in having applicable and useful deductions. These deductions help the researcher in presenting recommendations that contributes efficiently and effectively in improving the performance of studied companies and others with the same organization and techniques.

The study [5] titled “Validating A unified Framework for Knowledge Management” was a cadastal study performed with a number of researchers and managers in a variety of business sectors. The most important study outcomes were that knowledge management is linked to IT and electronic experience systems and is a knowledge field has its own social and economic philosophy, theory and basics. Also that it consists of many items such as leadership, technology and human element. The study confirms that it’s necessary to combine knowledge management in the academic domains in the universities.

3 RESEARCH PROBLEM STATEMENT

Through the experience of the researcher in the legal department in Saudi Arabian Airlines and the nature of this department which is full of expertise and depend largely on legal knowledge, this study tries to identify the degree of knowledge management practice in the department of legal affairs. The study problem is being tackled through answering the following query “What is the degree of practicing knowledge management operations in the department of legal affairs in Saudi Arabian Airlines?”

4 RESEARCH METHODOLOGY, TOOL AND DATA SAMPLE

The study used the descriptive approach “case study” to study the degree of knowledge management practice in the department of legal affairs in Saudi Arabian Airlines. The irregular interview is used and a questionnaire as a tool to collect data. The community of this study was the department of legal affairs section and its twenty employees. The sample would be the four heads of the four departments; investigations department, claims department, consulting department and contracts department and all heads have been met.

5 NUMERICAL FINDINGS

After the researcher has interviewed the deputy manager of the legal department and filled the questionnaire, he came to the following digital outcomes: Generally, the value of knowledge management practice in the department was 2,17 with percentage of 43,36% which is a low rate as described in the following Table 1:

Figure 1 shows the rates of knowledge management practice in the department of legal affairs as follows:

From Figure 1 and Table 1, we found that the degree of knowledge management practice in the legal department in Saudi Arabian is low as it is only 43,36%. The researcher would explain in detail the digital outcomes to each axis in knowledge management practice in the legal department in Saudi Arabian Airlines as follows:

Firstly, the department practice axis in the field of gaining knowledge.

The previous Table 2 shows that paragraphs varied between medium and almost low while this paragraph “The department director shares his experience in the administrative developed experiences of performance with his agents” has recorded the highest index with percentage of 80%. Whereas the paragraphs refers to attracting the experienced and academics has recorded the lowest indexes which is 20%. The axis in general has recorded 55, 71% with a medium index.

Secondly, the department practices in the field of generating knowledge.

Table 3 shows that the paragraphs varied almost between low and very low while the paragraph concerning the support of the department manager for the initiatives recorded the highest index with percentage of 80%. But paragraphs concerning organizing creativity and innovation competitions, cooperating with human resources and community institutions, attracting media means and bonuses recorded the lowest index with percentage of 20%. The index in general recorded low index with percentage of 38%.

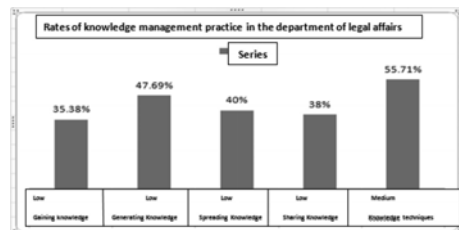


Figure 1. Rates of knowledge management practice in the department of legal affairs.

Table 1. Percentage and value rate of knowledge management practice in the legal department.

Axis	Index rate	Value rate	Percentage rate
Total percentage of management practice axis in the field of gaining knowledge	medium	2.79	55.71%
Total percentage of management practice axis in the field of generating knowledge	low	1.9	38%
Total percentage of management practice axis in the field of spreading knowledge	low	2	40%
Total percentage of management practice axis in the field of exchanging and sharing knowledge	low	2.39	47.69%
Total percentage of management practice axis in the field of knowledge techniques	low	1.77	35.38%
Percentage and value rate of knowledge management practice in the legal department	low	2.17	43.36%

Table 2. The practice of the legal department in gaining knowledge field.

Paragraph	Indexes and percentages		
	Index	Value	Percentage
Determining the employees' training needs	medium	3	60%
Setting programs and courses for professional development to employees	low	2	40%
Determining deficiencies of performance levels	medium	3	60%
Setting remedial plans to come over deficiencies	medium	3	60%
The department brings the experienced to give lectures to the employees	very low	1	20%
The department attracts the academic expertise in the field of business administration to develop the performance of the employees	very low	1	20%
The department director shares his experience in the administrative developed experiences of performance with his agents	high	4	80%
The department cooperates with training centers to select courses that fulfill the employees' training needs	medium	3	60%
The department manager cooperates with experts to determine the employees' knowledge requirements	medium	3	60%
The department manager shifts his gained knowledge from courses to the agents and employees	high	4	80%
The department facilitates the way for employees to get knowledge continuously	medium	3	60%
The department manager cooperates with the human resources manager in preparing employees to gain knowledge	medium	3	60%
The department manager organizes scientific trips for employees to enrich their knowledge and support performance	medium	3	60%
Including financial bonuses and charges related to gaining knowledge in the department budget	medium	3	60%
Total percentage of management practice axis in the field of gaining knowledge	medium	2.79	55.71%

Thirdly, department practice in the field of spreading knowledge.

Table 4 shows that the paragraphs varied almost between medium and low while this paragraph concerning using social media periodically and utilizing technological novelties recorded the highest index with percentage of 60%. But paragraphs

concerning with spreading knowledge and opening accounts in social media websites recorded the lowest index with percentage of 20%. The axis in general has recorded low index with percentage of 40%.

Fourthly, the department practice in the field of exchanging and sharing knowledge.

Table 3. Department practice in the field of knowledge generation.

Paragraph	Indexes and percentages		
	Index	Value	Percentage
The department manager supports the employees' initiatives in work field to develop performance	high	4	80%
The department manager encourages employees on scientific research	medium	3	60%
The department manager encourages performance on scientific creativity and innovation	low	2	40%
The department manager organizes research, creativity competitions and motivates the excellent with rewards	very low	1	20%
The department manager cooperates with human resources in adopting the performance of those who have excellent and innovative talent	very low	1	20%
Cooperation with community institutions in adopting the creative	very low	1	20%
Organizing courses that helps employees in creativity and innovation operations	low	2	40%
The department brings media means to do its role in supporting the creative	very low	1	20%
Building good example of employees as a creative symbol to the rest of their colleges	medium	3	60%
Including financial bonuses and charges related to gaining knowledge in the department budget	very low	1	20%
Total percentage of the department practice axis in the field of knowledge generation	low	1.9	38%

Table 4. Shows practices of the legal department in the field of spreading knowledge.

Paragraph	Indexes and percentages		
	Index	Value	Percentage
Periodical paper cultural bulletins	low	2	40%
Messages through social media periodically for promote the knowledge level	medium	3	60%
Sharing in meeting places and conferences related to work for spreading and attracting knowledge through them	low	2	40%
Utilizing journals, academic thesis and researches related to work in spreading knowledge	low	2	40%
Utilizing the department website in spreading knowledge	very low	1	20%
Utilizing technological novelties such as smart devices in spreading knowledge	medium	3	60%
Opening accounts in social media to spread knowledge	very low	1	20%
Holding cultural meetings in the department to publish creative and innovative related to the department	low	2	40%
Total percentage of the department practice in the field of spreading knowledge	low	2	40%

Table 5 the legal department practices in the field of exchanging and sharing knowledge as follows:

Table 5 shows that the paragraphs varied almost between medium and low while the paragraph "Meetings between department manager and employees and purposeful discussion, exchanging knowledge and experience" recorded the highest

index with percentage of 80% and the paragraphs concerning deepening community participation and organizing conversational programs periodically recorded the lowest index with percentage of 20%. The axis in general has recorded low index with percentage of 47, 69%.

Fifthly: the department practice in the field of knowledge techniques.

Table 5. The department practice in the field of exchanging and sharing knowledge.

Paragraph	Indexes and percentages		
	Index	Value	Percentage
Participating in purposeful meetings to exchange knowledge and experience	medium	3	60%
Encouraging employees on attending purposeful meetings to exchange knowledge and experience	medium	3	60%
Arranging exchangeable visits between employees to exchange experiences	medium	3	60%
Encouraging executives on exchanging experiences with their colleges in other departments to develop performance	low	2	40%
Cooperating with other department managers in the field of exchanging experience and knowledge	medium	3	60%
Cooperating with human resources in the field of exchanging experience and experiments	low	2	40%
Participating in the programs of exchangeable visits between departments locally and regionally	low	2	40%
Organizing lectures and seminars inside and outside the department	very low	1	20%
Encouraging employees on exchanging knowledge and experience	medium	3	60%
Encouraging department heads on exchanging experience with others in the field of practical problems and solutions	medium	3	60%
Deepening community participation in the field of exchanging experiences and professional and knowledge development	very low	1	20%
Organizing, recording and documenting conversational programs periodically	very low	1	20%
Meetings between department manager and employees and purposeful discussion, exchanging knowledge and experience	high	4	80%
Total percentage of department practice axis in the field of exchanging and sharing knowledge	low	2.39	47.69%

Table 6. The department practice in the field of knowledge techniques.

Paragraph	Indexes and percentages		
	Index	Value	Percentage
The department supports employees with techniques that helps them in spreading, gaining and exchanging knowledge	medium	3	60%
Building internal community communication network	high	4	80%
Creating a website for the department in a way that supports knowledge, spreading and gaining it	very low	1	20%
Providing methods of communication with information centers, data basis and digital libraries	low	2	40%
Investing information and knowledge techniques for organizing, storing and spreading knowledge	medium	3	60%
Supporting e-learning and increasing knowledge through it	low	2	40%
Supporting knowledge participation electronically through social media	low	2	40%
The department have an account in facebook for community communication and knowledge support	very low	1	20%
The department have an account in twitter for community communication and knowledge support	very low	1	20%
The department have an account in youtube for community communication and knowledge support	very low	1	20%
The ideal use of email in the field of spreading and exchanging knowledge	very low	1	20%
Linking the department website with electronic libraries and data basis	very low	1	20%
Supporting electronic training in the department using the internet	very low	1	20%
Total percentage of department practice axis in the field of knowledge techniques	low	1.77	35.38%

Table 6 shows the legal department practices in the field of knowledge techniques as follows:

Table 6 shows that the paragraphs varied almost between medium and low while the paragraph "Building internal community communication network" recorded the highest index in percentage of 80%. But the paragraphs concerning creating a website supporting participation, knowledge gaining and sharing and social media accounts recorded the lowest index with percentage of 20%. The axis in general recorded low index with percentage of 35, 38%.

6 SUMMARY, SUGGESTIONS AND RECOMMENDATIONS

The study came at that the degree of department practice in the field of gaining knowledge reached 55, 71% with medium index. Also, the degree of department practice in the field of knowledge generation reached 38% with low index while reached 40% with low index in spreading knowledge. The degree of exchanging and sharing knowledge reached 47, 69% with low index and the degree of knowledge techniques reached 35, and 38% with low index. Generally, the degree of knowledge management practice in the legal department reached 43, 36% with low index.

The study suggests that the department brings the experienced to give lectures for employees and attracts the academic experiences in the field of business administration to develop employees' performance. It also suggests organizing research and creativity competition and motivating the excellent with rewards. Cooperating with human resources in adopting the performance of the creative and innovative talent and with community institutions in adopting the creative would be useful in addition to attracting media means by the department to play its role in supporting the creative. The inclusion of financial bonuses and charges related to knowledge generation in the department budget would be of a

good benefit. For spreading knowledge, the department should utilize its website, open accounts in social media and organize lectures and seminars inside and outside the department itself. Additionally, the department should deepen community participation with the university in the field of experience exchange and knowledge and professional development. Organizing, recording and documenting conversational programs also would be very helpful. This study recommends making a suggestion of knowledge management in the legal department in Saudi Arabian Airlines and performing the same sort of studies on the rest of Saudi Arabian Airlines sections and departments. Also, it recommends performing a study on planning knowledge management in general in Saudi Arabian Airlines.

REFERENCES

- [1] Darwza, Suzan Saleh (2008). "Relationship between requirements knowledge management and its processes and on institutional performance, applicable study of Jordanian high education", unpublished master thesis, MEU for Advanced studies.
- [2] Abo El ela, Laila Muhammad Hosni (2012). "The degree of practicing knowledge management operations in the Education faculty in Taif from teaching staff's point of view". The international specialized educational journal. 1 (4) May 2012.
- [3] Abo khaseer K. Eman Saoud (2009). "Applications of knowledge management in high education institutions", Thoughts and practices. The international conference for administrative development (For excellent performance in the governmental sector). Institute of Public Administration. Riyadh 1-4, November 2009.
- [4] Al-fares, Soliman (2010). "The role of knowledge management in promoting efficiency of the organization performance (Field study on turning industries companies) in Damascus". Damascus journal for legal and economic sciences. 26(2).
- [5] Carolyn, B (2002) Validating Aunified Framework for knowledge management./index.htm02www./CASit.org/km/kmrt/may.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

A novel adaptive e-learning model matching educator-student learning styles based on machine learning

Manal Abdullah, Areej Y. Bayahya, Ebtesam S. Ba Shammakh,
Khawlah A. Altuwairqi & Areej A. Alsaadi

Department of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT: E-learning is becoming increasingly popular in many educational institutes. During the past decade, the researchers focused on the personalized e-learning systems and adaptive learning systems. This study proposed an adaptive model, Mat-ES, which acronym for Matched Educator-Student. It matches between educator's and student's learning style by using several classifiers such as Bayes, decision tree and SVM. In addition, it compares between students' performance using Mat-ES and the traditional way. The results show that the student's performance improved using Mat-ES. Also, J48 classifier gives the best results, because it had highest area under ROC and lowest MAE.

1 INTRODUCTION

In any educational environment, at any level, the teaching-learning process consists of three basic elements: a person who teaches (educator), a person who learns (students), and educational contents provided by educator [1, 2]. E-learning is becoming increasingly popular in many educational institutions. It ends up as the most important tool that any educational community will not raise up without having it. It has been taken place since over twenty years, and its application has started in the mid 1990 [3]. It allows the student accessing information without temporal or spatial restrictions. E-learning is different from traditional education where it does not require physical face-to-face interaction between educator and student to deliver educational contents [4]. Nowadays, E-learning becomes very comprehensive and universal. Therefore, traditional education systems could be replaced [2].

The educational system may detect learning styles of the learners and classifies users that interact with it. Learning style models over the past 25 years has brought increasing attention to the idea that students learn according to their cognition and preferences as stated by Hawk and Shah [5]. In addition, knowing the students' learning style leads to facilitate the educational process and increase students' performance.

E-learning uses different classification methods to: achieve students' patterns, create students' profiles and define learning path. Nearest neighbor, Naïve Bayes classifier, decision trees and neural networks are examples of common classifiers.

WEKA mining tools provide these classifiers and is used for extracting patterns from data.

An important aspect in success of any educational process is matching between the student's styles and educator's styles in teaching learning objects. A Learning Object (LO) [6] is defined as a material used many times in several courses or in different situations. Educators use LOs to help students to understand subject taught as well as raises learning motivation, interest and enjoyment among students. Most educators adopt a style of teaching which matches their own learning style, while it may be different from the student's style. Many mismatches can be avoided if both the student and educator take the time with high attention to complete the learning style questionnaire. Not only mismatches are avoided, but also awareness of learning style as an expansion of both learning and teaching styles may take place. This is critical since one of the roles of social specialist is to meet the students' individualized needs within inclusive special education program specially for young students such as KG1 and KG2.

Nowadays, lack of proper methods of teaching is not the only problem in the learning process. In fact, there are mismatching between student's style and educator's style. There are many reasons behind this problem. Sometimes educators do not put adequate efforts to find suitable learning objects. Another problem is that each educator defines its own methods in teaching process, which not derived from reliable academic sources. In personalized learning, the teaching method should be customized to students' style of learning.

This paper consists of five sections: Section II is literature review and related work. Mat-ES model and its functionalities is presented in section III. Section IV describes initial experiments. Results discussion and conclusions with future work will be explained in sections V and VI respectively.

2 LITERATURE REVIEW

2.1 *E-learning*

E-learning is a new teaching-learning way that supports the idea of gaining educational contents from anywhere at any time. Babic [2] defines e-learning as “an innovative approach to delivering electronically mediated, well designed, student-oriented, interactive e-learning environment, independent of the place and time, whether by using Internet or digital technologies according to the instructional design principles”.

Learning Management Systems (LMSs), such as WebCT, Moodle and Blackboard [7] are essential tools used by many educational institutions to support educator in creating and managing online courses [8]. During creating e-learning content, both courses and students’ profiles have to be stored in the database of the system. The course could be provided with a great several features such as learning material, quizzes, discussion forums, assignments, and so on [9].

Faculty members at King Abdulaziz University have used blackboard BB system since 2014 [10]. This Learning Management System (BB) [11] is used in this study for building the adaptive e-learning system. It is an integrated e-learning solution. It is used to follow the students’ performance and monitor the efficiency of the learning process. It also provides a safe and user-friendly learning environment. Educators carryout the courses and lectures via multimedia usage (text, images, audio, video, animation). Students come together to browse through the contents as per each individual’s needs. Students freely communicate with each other without restriction of time and place via the various communication tools (e-mail, forums... etc) or via the virtual classes which can operate from any smart device.

2.2 *Learning style*

A learning style term depicts the states of mind and practices, which define an individual’s preferred method for learning. Feldman et al. [12] states that “[a learning style is] the composite of characteristic cognitive, affective, and physiologi-

cal factors that serve as relatively stable indicators of how a student perceives, interacts with, and responds to the learning environment”. Its role is particularly important in educational settings since it helps students and educators become more self-aware of their strengths and weaknesses. It can be recognized by traditional way using a test or questionnaire. A large number of learning style models are proposed in the literature [5]. Felder and Silverman Learning Style Model as proposed in Ahmad and Shamsuddin [13] and Liyanage et al. [14], Kolb Experiential Learning Model as proposed in Kaninen [15], Felder and Silverman Learning Style Model, VARK Model as proposed in Hassan and Hegazy [16], Dunn and Dunn Model as proposed in Larkin-Hein and Budny [17], Gregorc Learning Style Model, and RASI Model.

In this study, researchers adopted two learning style models Kolb and VARK to identify the teacher’s and student’s learning style. Kolb for teaching process and VARK for learning process respectively. The reason behind the use of these models is that VARK [18] self-knowledge model finds out the preferred learning method to help students in learning and self-learning processes. As such, educators are required to increase the quality of teaching and learning process by creating a learning environment that attracts the attention of students to deliver the course according to the students’ needs. VARK enhances students’ understanding to the subject taught as well as raises learning motivation, interest and enjoyment among students. On the other side, Kolb’s experiential model [19] can help educators to reach the students of various mentalities in the classroom. Searching for pedagogical strategies is still the main concern for educators. Kolb’s found that educators could use the four stage cyclical process to design effective lessons. Educators can help students move from experience to observation to conceptualization to experimentation and then continue this cycle.

Neil Flemming [20], of Lincoln University, and his colleagues have developed VARK model. VARK classifies students based on four different modes. These modes are based on four different senses, namely visual, aural, reading, and kinesthetic, and the model name originated from those senses prefix letters (V, A, R, and K). Table 1 summarizes the characteristics of students based on these modes [5, 18, 20].

The Multimodal students are those students who have inclinations in more than one mode. The different possibilities of multimodal are presented in Table 2, where Binary are those with two modes, Tri with three modes and Quad with all Quadratic modes of VARK [16]. In this research, authors focused on the tri-possibilities VAK because (R)

Table 1. VARK model dimensions.

Visual (V)	Visual student gains the knowledge best by seeing. They like to use diagrams, graphs, charts, written text...etc.
Aural (A)	Aural student gains the knowledge best by listening. They learn best from discussion, audio tapes, Seminars...etc.
Read/Write (R)	Read/Write student prefers printed word and text as a method to gain information. They like to use textbooks, lecture notes, essay, handouts...etc.
Kinaesthetic (K)	Kinaesthetic student gains the knowledge best by doing, experience and practice. They prefer to apply touch, movement, and interaction to their environment.

Table 2. VARK “Multimodal Type” possibilities [16].

Bi possibilities	Tri possibilities	Quad possibilities
VA, VR, VK, AR, AK, RK	VAR, VAK, VRK, ARK	VARK

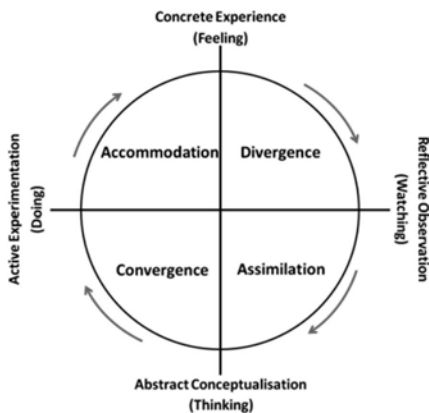


Figure 1. Kolb's learning style [15, 21].

Read/Write students are those who like lists, essays, reports, textbooks, definitions, printed handouts, readings, manuals, Web pages, dictionaries, thesaurus and taking notes. Therefore, (R) does not apply to programming course, where this research used as experimental course.

David Kolb [15, 21] developed Experiential Learning Theory (ELT) in 1984. Kolb's learning theory consists of four different learning styles, which are passing through a four-stage learning cycle, Figure 1. The cycle begins by a certain experience that makes the student eager to learn more

Table 3. Kolb learning style model dimensions.

Divergence	The diverging learning style is dominant by concrete experience and reflective observation. A student of this style is learning by feeling and watching.
Assimilation	The assimilating learning style is dominant by reflective observation and abstract conceptualisation. A student of this style is learning by watching and thinking.
Convergence	The converging learning style is dominant by abstract conceptualisation and active experimentation. A student of this style is learning by doing and thinking.
Accommodation	The accommodating learning style is dominant by active experimentation and concrete experience. A student of this style is learning by doing and feeling.

about this experience. For that to happen, student has to reflect on that experience from many different points of view as possible. From this reflection, student draws conclusions to make decision or take action. The characteristics of each style are summarized in Table 3.

2.3 Classification process

Classification is the task of identifying the objects in certain groups to their appropriate categories by building a model based on one or more numerical and/or categorical variables (predictors or attributes). The goal of classification is to be predictable for each data accurately and correctly [22].

Classification falls in two types: supervised and unsupervised. Supervised is used for classification purposes. It depends on selecting a training dataset to build a model capable to measure the performance and accuracy of testing dataset by using some classification algorithms. On the other hand, unsupervised is used for clustering purposes. It tries to find out a specific structure for an unlabeled dataset. In unsupervised no labeled training sets provided and the system applies a specified clustering or grouping to the unlabeled datasets based on some similarity criteria [22].

Ahmad and Shamsuddin [13] conducted comparative analysis of mining techniques using Naïve Bayes, Bayes Net, J48, NBTree, random-Tree, CART, Conjunctive Rule, Decision Table, and DTNB for automatic detection of student's

learning style. In this paper, researchers use the supervised classification algorithms such as J48 as Decision Tree algorithm [23], Naïve Bayes classifiers [24], NBtree [25], and SVM [26].

2.4 Adaptive learning system

Adaptive learning [27] is a system that has capability to modify its behaviors to provide different learning objects for every student based on his personalization characteristic like goal, knowledge, experience, background, interest, etc. Those features called user model, which are varying from a person to another, Figure 2. By using this adaptation technique, the application displays the most interesting information to the student, which leads to information spaces [28].

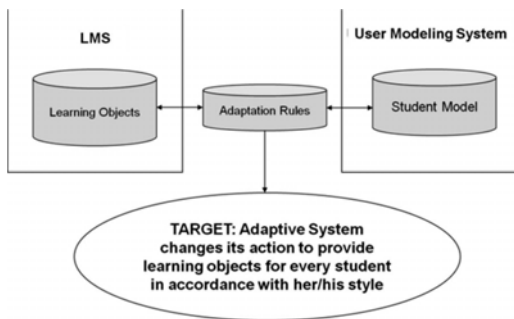


Figure 2. Interaction between user modelling system and LMS.

3 MATCHED EDUCATOR-STUDENT LEARNING STYLES MODEL

Matched Educator-Student learning styles (Mat-ES) is an adaptive learning system model proposed in this study. It depends on classifying students into different patterns to determine the suitable method of learning. However, it matches educator's style with student's style to maximize student's learning outcomes. Mat-ES passes in three phases: traditional learning, self-learning and personalized learning. The traditional learning is applied on all sections whereas self-learning and personalized learning are applied on only one section. The self-learning is when a student self learns by suitable learning objects, while personalized learning is an adaptive process relies on learning style and student's grades.

Figure 3 shows the architecture of Mat-ES model. It consists of four main components: Data collection and Pre-processing, Traditional Learning Model (TLM), Learning Management System-Black Board (LMS-BB) Model and Adaptation Model (AM). Next subsections detail these components.

3.1 Data collection and pre-processing

The researchers applied this study on students from King Abdulaziz University KAU, Jeddah, SA. The study was conducted on two sections of programming course, CP203. Mat-ES model has three different data sources: learning style classes

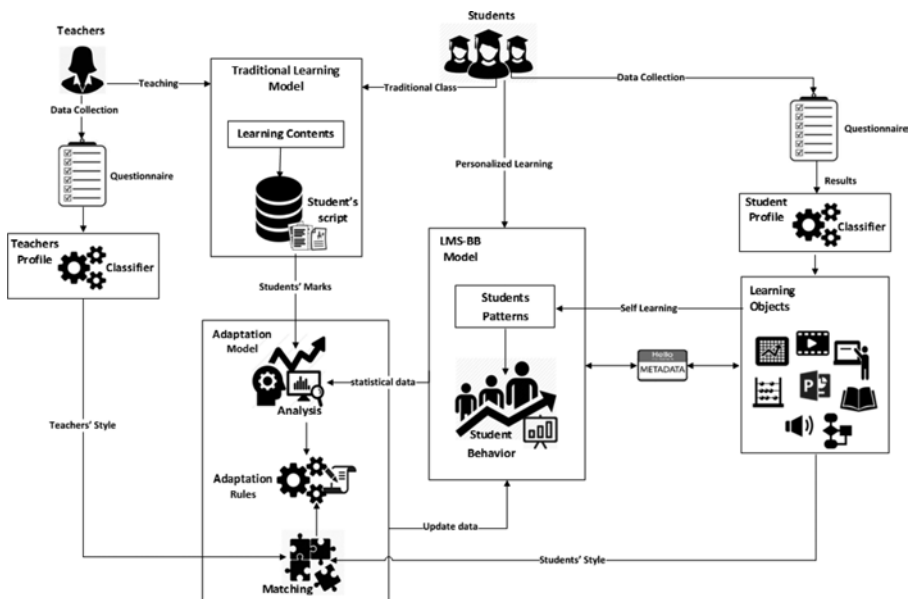


Figure 3. Matched Educator-Student learning styles (Mat-ES) mode.

of the students that is output of analyzing questionnaire developed by Kolb and VAK, grades, and students' behavior on BB LMS.

Student participates in the questionnaire, from which researchers generate students learning styles. Figure 4 shows sample of questionnaire output. This information is stored in the student's profile. In addition, the students need to log into LMS-BB to learn by their preferred learning objects, which is self-learning. The BB system contains all different learning objects.

3.2 Traditional Learning Model (TLM)

TLM is a traditional class, which based on face-to-face interaction between students and educators. Students take lectures and attend lab activities. TLM applied on students in both sections, while one section applied adaptive Mat-ES in addition. Consequently, Mat-ES compared and evaluated performance of students of both sections. TLM will send results after each 30% of student's marks as depicted in Figure 5. At phase 1, students study traditionally. Then, researchers apply adaptive Mat-ES on one section at phase 2 and up.

3.3 Learning management system-blackboard model

During phase 2, students start self-learning, by logging into BB and interact with all learning objects. In phase 3, students start personalized learning, which based on their preferred objects. The students is classified into different classes based on their learning style and behavior on BB. Each class mapped with its suitable learning

7: Total Numeric	8: Grade Nominal	9: VAK Nominal	10: Kolb Nominal	11: Learning Style Nominal
80.0	B	Visual	Convergence	Visual and Convergence
86.0	B+	Visual	Accommodation	Visual and Accommodation
61.0	D	Visual	Divergence	Visual and Divergence
82.0	B	Visual	Assimilation	Visual and Assimilation
75.0	C+	Kinesthetic	Convergence	Kinesthetic and Convergence
76.0	C+	Visual	Convergence	Visual and Convergence
77.0	C+	Visual	Accommodation	Visual and Accommodation
95.0	A+	Visual	Assimilation	Visual and Assimilation
85.0	B+	Visual	Convergence	Visual and Convergence
93.0	A	Visual	Assimilation	Visual and Assimilation

Figure 4. Sample of classification learning style.

30%						Total grade	30%				Total grade	40%	Final Learning Object	
Quiz 1	Quiz 2	Quiz 3	Final Midterm	Quiz 4	Quiz 5	30	Quiz 6	Quiz 7	Second Midterm	Assignment	Total grade	Final Exam	Learning Object	
25	3	25	15	11	25	3	3	125	575	2575	19	19	Slide	
3	25	3	15	125	25	15	265	3	3	1425	590	2670	20	Diagram
3	3	3	12	3	3	27	3	3	125	590	2730	20	Table	
1	25	3	3	175	3	3	2625	3	2	11	590	2180	20	Table
2	3	3	25	12	25	2	27	25	3	125	480	2580	22	Table

Figure 5. Sample of students' grades.

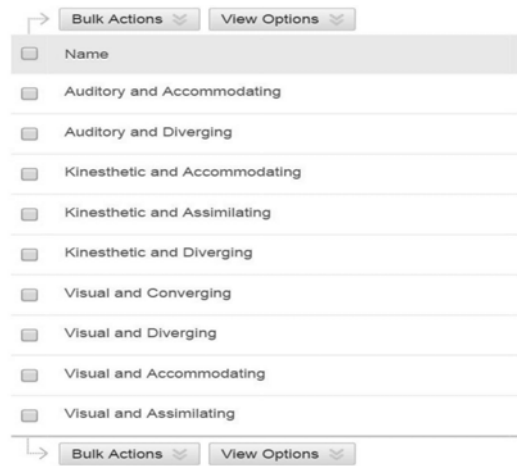


Figure 6. Students' classes.

```

VAK = Aural: Aural and Accommodation (3.0/2.0)
VAK = Kinesthetic
| Kolb = Accommodation: Kinesthetic and Accommodation (2.0)
| Kolb = Assimilation: Kinesthetic and Accommodation (0.0)
| Kolb = Divergence: Kinesthetic and Divergence (2.0)
| Kolb = Convergence: Kinesthetic and Convergence (2.0)
VAK = Visual
| Kolb = Accommodation: Visual and Accommodation (7.0)
| Kolb = Assimilation: Visual and Assimilation (13.0)
| Kolb = Divergence: Visual and Divergence (9.0)
| Kolb = Convergence: Visual and Convergence (8.0)

```

Figure 7. Adaptation rules.

objects. Those classes stored on BB as students' patterns. Figure 6.

3.4 Adaptation Model

AM receives statistical reports and student's grade from BB and TLM respectively. Based on this input, system analyzes the received data and assigns the adaptation rules, see Figure 7. Moreover, this model applied adaptation rules between educator's style and student's style. Finally, AM will repeat adaptation rules until reached the best learning objects. After each phase, system checks the student performance. LO is changed if and only if the student's performance decreased. So, student's grades are the main factor to replace or keep on learning object on BB.

4 EXPERIMENTS

4.1 Experimental setup

The students of Programming course have been chosen because the programming courses are

considered the main pillar for computer science students. The participants were second-year undergraduate students collected from two sections, named as section A and section B, respectively, in the Department of Computer Science and Information technology at KAU in Saudi Arabia. 46 students and 2 educators have filled the questioner. Section A has been taught traditionally. On the other hand, section B has been taught based on the Mat-ES model (Adaptive and self-learning).

There are 3 applied phases in section B. The first-phase has been done traditionally without any improvement. The Mat-ES sent students marks to AM (Adaptation Model) for later use. Then, second-phase starts by uploading all learning objects into BB. During this phase, students start self-learning. They need to logging into BB and interact with objects. The BB will browse all learning objects to all students without any restriction. In addition, students' behavior is registered in BB. The Mat-ES sent student's marks and Students' behavior to AM too. In the third-phase, the students were classified into different classes based on their learning styles obtained by questionnaire and their behavior collected from the previous phase. The suitable learning objects for each student are uploaded into BB, depending on their personalized. The Mat-ES sent student's marks and behavior to AM to apply the adaptive rules. At the end, Mat-ES gets the best student's class based on final grade. It analyzes the results to show if students' class compatible with educator's style or not. We applied this way in both Sections A and B. In addition,

we studied impact of adaptive method in case student's class incompatible with educator's style in section B. In Table 4, the second column shows the suitable learning objects based on VAK and Kolb learning styles while the used objects are displayed in the last columns.

4.2 Classification and pattern discovery

Analyzing the results of questionnaire is done to determine the preferred style for every student. In VAK style, the suitable class is the max value between the visual, auditory and kinesthetic. On the other hand, the appropriate style in Kolb is determining based on both max values of x-axis (RO and AE) and y-axis (CE and AC). It is the combination of two preferred styles such as Diverging and Assimilating. The students' classifications were determined based on (student's grade, VAK, Kolb). However, the student's style name was recognized by the dominated class of both VAK and Kolb i.e. Visual and Assimilating. Therefore, 12 different combinations of both styles are displayed in Table 4. In this study, only 10 classes of them are applied based on the questioner.

WEKA conducts the classification by built-in classifiers algorithms to train and test the dataset. WEKA contains many tools for data pre-processing, clustering, classification, association rules, regression, and visualization.

In this research, the dataset has been running on Naïve Bayes classifier, pruning algorithms (J48, NBTree) of classification tree, and SMO to

Table 4. The classes with the suitable and applied materials [5].

The class	The suitable materials	The applied material
Visual and Diverging	Graphics, Tables, Images, videos, Written instructions, A brainstorming session, Animation, Diagram	✓ Diagram
Visual and Assimilating	Graphics, Tables, Images, Videos, Written instructions Audio, video, lectures, verbal tutorials	✓ Diagram ✓ Video
Visual and Converging	Graphics, Tables, Images, Videos, Written instructions, Text-based materials such as Microsoft office (PowerPoint, Word, Excel), web pages	✓ Diagram ✓ Slide ✓ Web pages
Visual and Accommodating	Graphics, Tables Images, Videos, Written instructions, Web pages	✓ Diagram ✓ Web pages
Auditory and Diverging	Discussion, Cooperative learning groups, podcast, A brainstorming session, Animation, Diagram	✓ Diagram ✓ Discussion
Auditory and Assimilating	Discussion, Cooperative learning groups, podcast,	
Auditory and Converging	Discussion, Cooperative learning groups, podcast,	
Auditory and Accommodating	Discussion, Cooperative learning groups broadcast, web pages	✓ Web pages ✓ Discussion
Kinesthetic and Diverging	Playing games, Read body language, A brainstorming session, Animation, Diagram	✓ Diagram
Kinesthetic and Assimilating	Playing games, Read body language, Video	✓ Video
Kinesthetic and Converging	Playing games, Read body language,	
Kinesthetic and Accommodating	Playing games, Read body language, Web pages	✓ Web pages

increase the speed of the training algorithm. All of those are applied in WEKA data mining tool. In both algorithms namely classification tree (J48, NB-tree) have selected with pruning (Naïve Bayes) and SVM decision boundaries algorithms (SMO tools) Classifier.

Naive Bayes classifiers are one of the old and well-known classifiers. They are using the probabilistic approach. They try to calculate conditional class probabilities and then predict the most likely class.

NBTree is a hybrid algorithm between Decision Tree and Naive-Bayes. In specific, decision tree with leaf nodes of naive Bayes classifiers.

The main idea behind SVM is the construction of an optimal hyperplane, which can be used for classification, for linearly separable patterns. The optimal hyperplane is a hyperplane selected from the set of hyperplanes for classifying patterns that maximizes the margin of the hyperplane i.e. the distance from the hyperplane to the nearest point of each patterns. The main objective of SVM is to maximize the margin so that it can correctly classify the given patterns i.e. larger the margin size more correctly it classifies the patterns.

4.3 Model validation

The applied classifiers are evaluated by using 2 test modes which are the percentage split and the 10-fold cross validation procedures. The first mode means classification results will be evaluated on a test set. The data is split into training and test set. It takes the given percentage of the data and trains the classifier using that percentage. On the other hand, a 10-fold cross validation procedures are 10 folds (parts), afford each part in turn, and averages the results. Each point of data is utilized for testing only one time and nine times for training. At the end of both modes, the results of the experiment were compared in terms of students' performance.

5 RESULTS AND DISCUSSION

As mentioned in the previous section, there are two different test modes applied on this experiment: Percentage split and 10-fold cross validation. Percentage split is the first test-mode applied. In this mode, four different training-testing dataset are applied: (40%-60%), (60%-40%), (65%-35%) and (75%-25%). The split ratio (65%-35%) gives the highest accuracy as shown in Table 5. The second test-mode is cross-validation. In this mode, dataset is divided into 10 parts (folds), hold out each part in turn and take the average of all result. Each part is used once for testing and 9 times for training. Table 6 shows test-mode (10-Fold

Table 5. Test mode (split 65% train, 35% test).

Classifier algorithm tools	Correctly classified instances (%)	Kappa statistic	MAE (%)	Weighted Avg.		
				TP	FP	ROC
J48	100	1	0	1	0	1
NBTree	100	1	0.0614	1	0	1
Naïve Bayes	75	0.6578	0.1029	0.75	0.71	0.748
Naïve Bayes + SMO	81.25	0.7348	0.0375	0.813	0.071	0.871
SMO	100	1	0.16	1	0	1

Table 6. Test mode (10-fold cross-validation).

Classifier algorithm tools	Correctly classified instances (%)	Kappa statistic	MAE (%)	Weighted Avg.		
				TP	FP	ROC
J48	89.1304	0.8677	0.0217	0.891	0.003	0.983
NBTree	89.1304	0.8654	0.075	0.891	0.021	0.979
Naïve Bayes	34.7826	0.2537	0.1381	0.348	0.063	0.66
Naïve Bayes + SMO	50	0.4019	0.2874	0.5	0.07	0.519
SMO	80.4348	0.753	0.1637	0.804	0.045	0.908

Cross-Validation). J48, NBTree, Naïve Bayes, improved Naïve Bayes (Naïve Bayes + SMO) and SMO classifiers are the applied algorithms in both test-modes. Various threshold was used in this dataset and the high accuracy appeared when setting threshold was 75 in final grades. This paper focus on percentage split of test modes, because it gives the highest accuracy.

Table 5 and Table 6 summarize the results as correct and incorrect classified instances. While accuracy measured by: Kappa statistic, Mean Absolute Error (MAE) and the weighted averages of True Positive rate (TP), False Positive rate (FP) and a Receiver Operating Characteristic (ROC) area.

Kappa statistic defined as frequently used to test inter rate reliability, which measured accuracy of final LS. Two or more independent observers are evaluating the same thing where 1 is perfect test, 0 is exactly what would be expected, and negative values indicate worthless test.

MAE measured accuracy differently than Kappa, which defined as a quantity used to measure how close predicted LS to a proposed LS. The smaller value indicate that the predicted LS closer to the proposed LS and has a high accuracy.

ROC is a graphical plot explain the performance of a binary classifier with different thresholds. The curve plots show the true positive rate against the false positive rate at various threshold settings. An area of 1 represents a perfect test.

In Table 5, Kappa and MAE of J48, NBTree and SMO measure performance of the final learning style classification. It had an accuracy of 100% with the value of kappa 1 and the value of MAE 0, 0.0614 and 0.16, respectively. All classifiers had higher accuracy than Naïve Bayes and improved Naïve Bayes. This percentage discovered that no error in J48, while the NBTree and SMO have 0.0614 and 0.16 error, respectively. Among these classifiers, J48 has the highest weighted average ROC, 1. Algorithms with highest area under the ROC and lower MAE have more powerful classification capability, hence j48 is the preferred algorithm for use. In Fig. 8, the Bar chart of Correctly Classified Instances in percentage split test mode show accuracy of all classifier algorithms. Fig. 9 show the accuracy measures rate of classifier algorithms in percentage split test mode. Fig. 10 show Area under ROC.

In Table 6, Kappa and MAE of J48 and NBTree measure performance of the final learning style classification. It had an accuracy of 89.1304% with the value of kappa 0.8677 and 0.8654, and the value of MAE 0.0217 and 0.075, respectively. J48 classifier had the highest accuracy and the highest weighted average ROC, 0.983. Consequently, J48 is the best algorithm for use in Percentage Split and Cross Validation test modes because it had the highest area under the ROC and lower MAE.

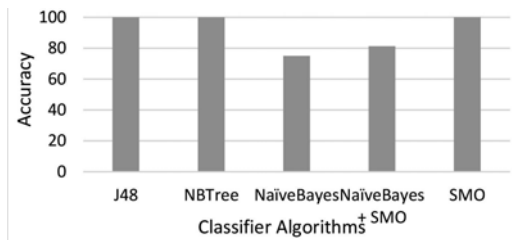


Figure 8. Correctly classified instances.

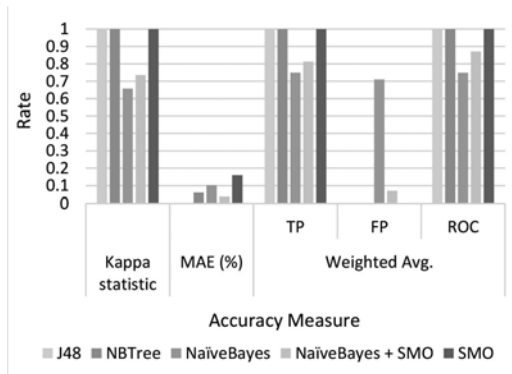


Figure 9. Accuracy measure rate.

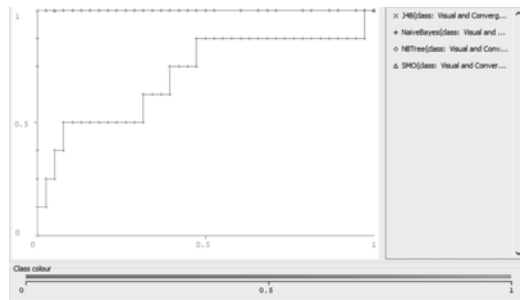


Figure 10. Area under ROC.

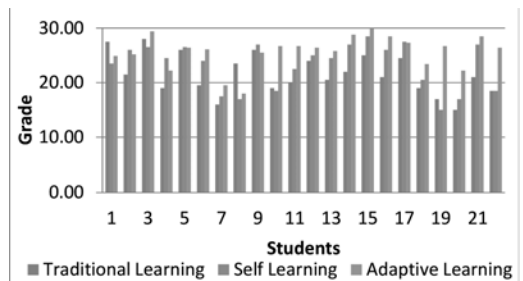


Figure 11. Students' grades in three different learning techniques.

In section B was applied three different Learning Techniques, Traditional Learning, Self-Learning and Adaptive Learning. Depending on the study and hypothesis, Adaptive Learning gave an improvement in the performance of students than Self-Learning and Traditional Learning as show in Fig. 11. The one-way analysis of variance (ANOVA) is used to determine whether there are any significant differences between three or more independent (unrelated) groups. If Alpha $\alpha \leq$ P-value the differences between some of the means are statistically significant. Otherwise, the differences between the means are not statistically significant. In first case, the Alpha value was selected $\alpha = 0.05$, P-value between In Traditional Learning and Self-Learning is $4.57638E-05$. In second case, P-value between Self-Learning and Adaptive Learning is $3.82485E-05$. In two cases $\alpha <$ P-value, but P-value in second case lower than P-value in first case. From the previous statistical Adaptive Learning improved students' performance than others.

In Section A, it was taking the average of final students' grades in each classes. The students' grades with Kinesthetic and Convergence learning style got the highest grades. Which they compatible with educator learning style.

In Section B, also it was taking the average of final students' grades in each classes. The students'

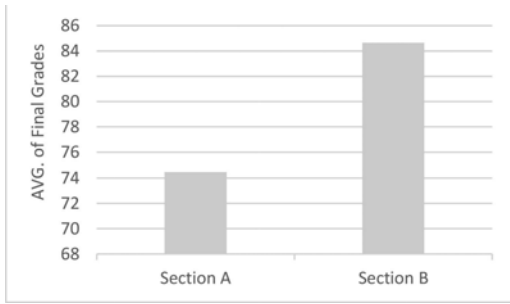


Figure 12. Compare between final grades in sections.

grades with (Kinesthetic and Divergence) learning style and (Aural and Accommodation) learning style got the highest grades. The students in class Kinesthetic and Divergence got the highest grades because educator learning style of section B compatible with them. While the students in class Aural and Accommodation are incompatible with their educator but they got the highest grades, because they have used adaptive learning techniques by used a lot of learning objects in BB many times.

In Fig. 12, the final grades of students in section A lower than the final grades of students in section B. Section A taught traditionally and Section B is taught based on Adaptive Learning. The different of students' grades between Section A and B based on way of taught. Each Students preferred specific method for learning, it's improved their performance.

6 CONCLUSION WITH FUTURE WORK

This study compared between the final grades of two sections of programming course. One of sections has been taught traditionally. The adaptive learning was applied to teach the other section. The adaptive learning is based on the proposed model which is Mat-ES to match between educator's and student's learning style. The applied learning styles are Kolb for teaching process and VARK for learning process.

The experimental results showed great contrast between students' final grades of both sections. The adaptive learning has improved the students' comprehension. The compatibility between the student's learning style and educator's learning style improves the student's grade in both types of learning (traditional and adaptive learning).

In future work, this study will be extended to apply onto more educators in programming course because the researchers want to get more accurate effect of the proposed module (this module was applied onto only two educators). Furthermore,

more learning objects are planned to add into LMS which gives more choices for students.

REFERENCES

- [1] H. Sarjono, S. Candra, and N. J. Setiadi, "From traditional learning into e-learning: Comparing students response to promote e-learning in college education," *Teaching, Assessment and Learning for Engineering (TALE), 2013 IEEE International Conference on*, pp. 7–11, 26–29 Aug. 2013.
- [2] S. Babic, "E-learning environment compared to traditional classroom," *MIPRO, 2011 Proceedings of the 34th International Convention*, pp. 1299–1304, 23–27 May 2011.
- [3] L. Zekanovic-Korona, V. Mateljan, and B. Krce Miocic, "Evaluation system for e-learning," in *MIPRO, 2012 Proceedings of the 35th International Convention*, 2012, pp. 1372–1376.
- [4] L. G. Muradkhanli, "Blended learning: The integration of traditional learning and eLearning," *Application of Information and Communication Technologies (AICT), 2011 5th International Conference on*, pp. 1–4, 12–14 Oct. 2011.
- [5] T. F. Hawk and A. J. Shah, "Using learning style instruments to enhance student learning," *Decision Sciences Journal of Innovative Education*, vol. 5, pp. 1–19, 2007.
- [6] J. M. C. da Silva and R. A. Silveira, "The Development of Intelligent Learning Objects with an Ontology based on SCORM Standard," in *Intelligent Systems Design and Applications, 2007. ISDA 2007. Seventh International Conference on*, 2007, pp. 211–216.
- [7] S. N. Hamade, "Student Perceptions of Learning Management Systems in a University Environment: Yahoo Groups vs Blackboard," in *Information Technology: New Generations (ITNG), 2012 Ninth International Conference on*, 2012, pp. 594–599.
- [8] I. Ruano Ruano, J. Garcia Gamez, and J. Gomez Ortega, "Building SCORM embedded WebLabs with LMS interaction," in *Frontiers in Education Conference (FIE), 2014 IEEE*, 2014, pp. 1–4.
- [9] S. Graf, Kinshuk, and L. Tzu-Chien, "Identifying Learning Styles in Learning Management Systems by Using Indications from Students' Behaviour," in *Advanced Learning Technologies, 2008. ICALT '08. Eighth IEEE International Conference on*, 2008, pp. 482–486.
- [10] King Abdulaziz University (2014) Blackboard system officially approved. Available at: <http://www.kau.edu.sa/Pages-Blackboard-system.aspx> (Accessed: 7 December 2015).
- [11] Deanship of e-Learning and Distance Education-King Abdulaziz University (2014) The learning management system blackboard. Available at: <http://elearning.kau.edu.sa/Pages-blackboard-info-e.aspx> (Accessed: 7 December 2015).
- [12] J. Feldman, A. Monteserin, and A. Amandi, "Automatic detection of learning styles: state of the art," *Artificial Intelligence Review*, pp. 1–30, 2014.
- [13] N. B. H. Ahmad and S. M. Shamsuddin, "A comparative analysis of mining techniques for automatic detection of student's learning style," in *Intelligent*

- Systems Design and Applications (ISDA), 2010 10th International Conference on*, 2010, pp. 877–882.
- [14] M. P. Pitigala Liyanage, K. S. L. Gunawardena, and M. Hirakawa, “A framework for adaptive learning management systems using learning styles,” in *Advances in ICT for Emerging Regions (ICTer), 2013 International Conference on*, 2013, pp. 261–265.
- [15] E. Kanninen, “Learning styles and e-learning,” *Master of Science Thesis, Tampere University of Technology*, 2008.
- [16] S. Hassan and A. El Fattah Hegazy, “A model recommends best machine learning algorithm to classify learners based on their interactivity with moodle,” in *Computing Technology and Information Management (ICCTIM), 2015 Second International Conference on*, 2015, pp. 49–54.
- [17] T. Larkin-Hein and D. D. Budny, “Research on learning style: applications in the physics and engineering classrooms,” *Education, IEEE Transactions on*, vol. 44, pp. 276–281, 2001.
- [18] N. Othman and M. H. Amiruddin, “Different perspectives of learning styles from VARK model,” *Procedia-Social and Behavioral Sciences*, vol. 7, pp. 652–660, 2010.
- [19] J. M. Baker, *Exploring technological literacy: Middle school teachers’ perspectives*: ProQuest, 2008.
- [20] A. M. H. Alawawdeh, A. S. Imran, and S. J. Kowalski, “Norwegian information security lectures as a case study for hyper interactive presenter,” in *Computer Applications and Information Systems (WCCAIS), 2014 World Congress on*, 2014, pp. 1–5.
- [21] J. Heywood, “Learning Strategies and Learning Styles,” in *Engineering Education: Research and Development in Curriculum and Instruction*, ed: Wiley-IEEE Press, 2005, pp. 119–151.
- [22] R. Sathya and A. Abraham, “Comparison of supervised and unsupervised learning algorithms for pattern classification,” *Int J Adv Res Artificial Intell*, vol. 2, pp. 34–38, 2013.
- [23] G. Kaur and A. Chhabra, “Improved J48 Classification Algorithm for the Prediction of Diabetes,” *International Journal of Computer Applications*, vol. 98, 2014.
- [24] A. Nurnberger, C. Borgelt, and A. Klose, “Improving naive Bayes classifiers using neuro-fuzzy learning,” in *Neural Information Processing, 1999. Proceedings. ICONIP ’99. 6th International Conference on*, 1999, pp. 154–159 vol.1.
- [25] P. Pumpuang, A. Srivihok, and P. Praneetpolgrang, “Comparisons of classifier algorithms: Bayesian network, C4.5, decision forest and NBTree for Course Registration Planning model of undergraduate students,” in *Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on*, 2008, pp. 3647–3651.
- [26] N. Satyanarayana, C. Ramalingaswamy, and Y. Ramadevi, “Survey of Classification Techniques in Data Mining.”
- [27] “Evaluating adaptive learning model,” in *Interactive Collaborative Learning (ICL), 2014 International Conference on*, 2014, pp. 818–822.
- [28] B. Qiu and W. Zhao, “Student Model in Adaptive Learning System Based on Semantic Web,” in *Education Technology and Computer Science, 2009. ETCS ’09. First International Workshop on*, 2009, pp. 909–913.

Feasibility for a seamless integration of admission, registration and academic advising in KAU-SIS

Fekry Fouad

King Abdul Aziz University, Jeddah, Makkah, Saudi Arabia

ABSTRACT: It is the vital role for the activities of Admission and Registration in King Abdul Aziz University to activate the main links between the university student information system from admission until graduation, through many tasks done by the deanship of admission and registration for undergraduate and graduate students. The concept of changing the mode of Information Technology (IT) of course would allow the system users and beneficiaries to make best use of utilizing databases and applications for Student Information System including all details of the Academic Advising through the current Student Information System (KAU-SIS) “King Abdul Aziz University Student Information System”, to have a successful the access of their records in a centralized mode. The most important changes is to run the online applications. This paper attempts—using SWOT analysis—to study the feasibility of a seamless integration of academic advising and KAU-SIS using the methodology to develop and evaluate a web-based Technologies for the Seamless Integration of Admission, Registration and Academic Advising in King Abdul Aziz University.

Keywords: student information system, online advising, Academic Advising, integration, Online Advisor, SWOT Analysis

1 INTRODUCTION

The integration of the Student Information System and the Academic Advising information supports all the academy staff as advisors and students to have best use of the current university student information system. KAU-SIS always provides information about education affairs researches as well as the scientific cooperation availability, and further education capabilities. The use of information technology, through KAU SIS is recognized historically as a legacy system platforms, with a slight superficial integration and an insufficient support of the workflow of business processes in KAU university, mostly administration level only. IT systems implement some programs that are oriented to certain administration concept.

The ongoing reforms of higher education and the adoption of degrees affect the further development of university software. New services are introduced such as academic advising services. The need for integration systems from increasing requirements to combine data throughout the whole university or department and to extract information for the university management.

At KAU, it is clear that advising services should be an integral part of the educational experience that we provide to our students. The real goal of the paper is to quantify and evaluate the online or web-based

student information system of KAU including the academic advising functionalities based on an integrated data. In order to get the research purpose, the following specific milestones were identified:

- a. To investigate the needed information of the existing student information system and the Academic Advising information.
- b. To study and identify the problems occurred in the current information system of students.
- c. To investigate all required information supports online mode.
- d. To study the security/control that are needed to run a safe online information system for students.
- e. To analyze the level of understanding of online system by:
 - The office of management/administration affairs;
 - The office of the Academic affairs;
 - The Stakeholders; and
 - The Students.

2 MAIN FOCUS

Still some universities as KAU use a semi—computer application as a core of their system only, same time they are completing their works

using a paper-based workflow especially for student information system. Always the staff feel tediousness even when they tried to prepare one report related to student's information and also exhausted of repetition and duplication of some work processes always to be done in updating and filling of records. The real objective of the university is to promote the sustainable resource to melt the student data and information with the university integrated information system.

By merging all the data of Academic Advising along with the university global data and use the new technique of the record management that will case unique information system.

Scope and delimitation

The recommendation to get an integrated Academic Advising and SIS in a unique university website will support all student services as a portal to be used by academic staff as well as student efficiently.

The system simplify all students, the recording and changing/updating of their personal sheet's, curriculums, and records of grades. The system also shows every student status either inactive or active. Especially for an inactive student status, it needs further orientation decision to transfer students or stopped or dropped subjects. The system also supports the monitoring and observing of different types of queries such as the students subjects to be enrolled.

As a prove of integration of the academic data, the report generation includes students' personal data sheet (that always given by the existing KAU-SIS) and curriculum details, and grade, It includes a integrated detailed and summarized report of inactive as well as active students, classified and categorized based on the academic details that given by both SIS and Academic Advising system.

Sources of data

Data used to be gathered by the Deanship of Admission and Registration (DAR) particularly in the registrar offices of the university. The registration staff were gathering of full information that was given by the students in the KAU DAR website.

3 REVIEW OF RELATED STUDIES

S.Marrero (2009) "Student Information System for the Universities" explained that the idea of Information Systems (IS) that come to the light in the early 1960s. Specifically, when trying to define "information system", as a part of Information

Science that always correlated to IS as an academic field (**emphasis academic**)

MIT (2009) proved that (SIS) Student Information System gives a comprehensive tool for the students to access their biographic data and academic (**emphasis academic**) to change or update their detailed information and pre-registration for different classes.

Evangelista (2008) shows that the (SIS) is a trusted, web integrated and online system that allows user to see their details.

Desousa (2008), confirmed that the Web based SIS have four core functions. 1) Harmonization of data elements, 2) Efficiency. That the users avoid to deal with paper transactions.

Swartz (2007) confirm that most of higher education institutions work in an integrated data. So we are preparing to a change in the method to gather, analysis and arrange/manage their information.

Principles of academic advising into KAU-SIS integration

Advising is a process that supports the relationship between the advisor and the advisee based on student Academic and Bio data availability. There are two different models of advising—prescriptive and developmental. Prescriptive advising entails the provision of information or explanation for a specific course of action. Developmental advising guides its practice on the premise that students are diverse and are at different stages of cognitive, interpersonal and psychosocial development. The advisor guides the student to explore options and to use available resources to become a more effective decision-maker based on identified learning outcomes. Developmental advising represents a learning-centered advising paradigm.

KAU, advising services encompass a developmental perspective but also a responsibility for service provision to all those key individuals and services that offer advice and assistance to students so as to enrich or improve their academic and social experience.

In some advising circumstances, such as learning about university regulations or course selection, prescriptive advice is appropriate. However, over the course of a student's academic career, the developmental advising model results in a more motivated, self-aware and successful individual.

Successful acquisition of learning outcomes furthers the university's quest for academic excellence, increases retention through a student culture of engagement, exploration and activism, enriches the student experience, and prepares graduates to take their places in the world as committed citizens, effective decision-makers, and engaged stewards and custodians of both community and environment.

4 ANALYZING ADVISING SERVICES AT KAU

To achieve its objectives for advising services at KAU, the Deanship of Admission and Registration (DAR) Student Support Services Center, first engaged the expertise of a consulting services to investigate the required data/information and functionalities that highly needed to be integrated with the KAU-SIS, to:

- Improve the way we respond to students in terms of quality of interaction and timelines;
- Focus on best practices;
- Standardize service levels at KAU; and
- Identify efficiencies and implement processes for improved access to services, monitoring of student requests and monitoring of performance.

The next step in the process, was to systematically examine the diverse advising services currently offered at the university. This included analyzing points of service, roles and responsibilities, sources of student, staff and management data/information and concerns, processes for tracking, and strengths/weaknesses and opportunities.

The outlined specific activities to help build a complete picture- and in-depth understanding of advising at KAU. These included:

- a. Two brainstorming sessions resulting in a SWOT Analysis;
- b. Focus groups with students, faculty members and staff on advising services; and
- c. Agreement with NACADA on an advising mandate and new vision.

Existing academic advising services

The Advising Cross-functional Committee first came together to provide a comprehensive description of all the steps and processes that a student passes through at KAU, from recruitment through to graduation, and all the services that “touch” students at the various stages in the cycle. Each of these phases or processes was then evaluated for efficiency and need for improvement. From this, three areas of particular focus were identified and prioritized intervention advising and outreach, developmental advising, and program specific advising. A list of all the units on campus delivering these types of services was then drawn up.

Defining the needs of the students

The next step in the advising project consisted of a brainstorming sessions to examine all aspects of advising services that presently take place on campus.

Advising services were assessed from a student perspective—defining student needs and expectations with regards to services, examining the accessibility of services (where do students go for help, how do they get access, who do they contact), and rating the current provision of services.

Over the multi sessions, the Working Group identified the main areas/functionalities of integration for which Academic Advising/KAU-SIS requires. These included:

1. Course selection
2. Program requirements
3. Rules and regulations
4. Program options
5. Academic standing

5 STRENGTHS WEAKNESS OPPORTUNITIES THREATS (SWOT) ANALYSIS

With the information from the SWOT Analysis and the data collected during the five different focus groups, the Working Group then identified common themes within consistent areas of concern, and began to examine opportunities and solutions that could help rectify the need of a seamless integration for the Academic Advising and KAU- SIS.

Mapping SWOT analysis results

The presentation SWOT analysis results in a graphical map gives a quick understanding of the main characteristics that highlights the results. The SWOT Mapping model show the differences between weakness and strengths in a different scales, and its strategic impact the makes relevance. The probability make a difference of opportunities and threats.

The X-axis

Any SWOT factor, does not show a strength or a weakness. The level of acceptance for KAU SIS or a strong KAU SIS could charted and mapped on axis with:

- Significance Strengthens 5.0
- Strengthen 3.0
- Minority of the Strengthens 2.0
- Neutrality of the results 0.0
- Weakness of the results -3.0
- Significances of the Weakness -5.0
- Minority of the Weakness -1.0

Accordingly, threats shows a negative opportunities, same time the threats and the opportunities shows also mapped on an on going axis with:

- Significance Opportunity 5.0
- Minority of Opportunity 1.0
- Opportunities 3.0
- Neutrality factors 0.0
- Minority of Threats -1.0
- Threat -3.0
- Significance of Threats -5.0

Then, (x-axis) shows the measuring of the SWOT analysis scale factor.

Y-axis

We can notice that the (strengths and/or weaknesses) do not prove the relevancy or importance's. It confirms the accepting of KAU SIS cause big difference in a way KAU SIS in a need for long time plan to enhance and generate a positive acceptance.

Accordingly, the opportunities or even the threats are not shown equality. The threats that make difference of occurring should be managed differently.

So, it is more convenient to use the probability as an axis when mapping threats and opportunities.

(Y-axis) is to map the relevancy of strengths and/or weaknesses and the most probabilities of threats or opportunities as follows:

- High Relevancy Strengthens/Weakness 5.0
- Relevancy Strengthens/Weakness 3.0
- Little Relevancy Strengthens/Weakness 1.0
- Neutrality 0.0
- Low Probability Opportunities/Threat -1.0
- Medium Probabilities Opportunities/Threats -3.0
- High Probabilities Opportunities/Threats -5.0

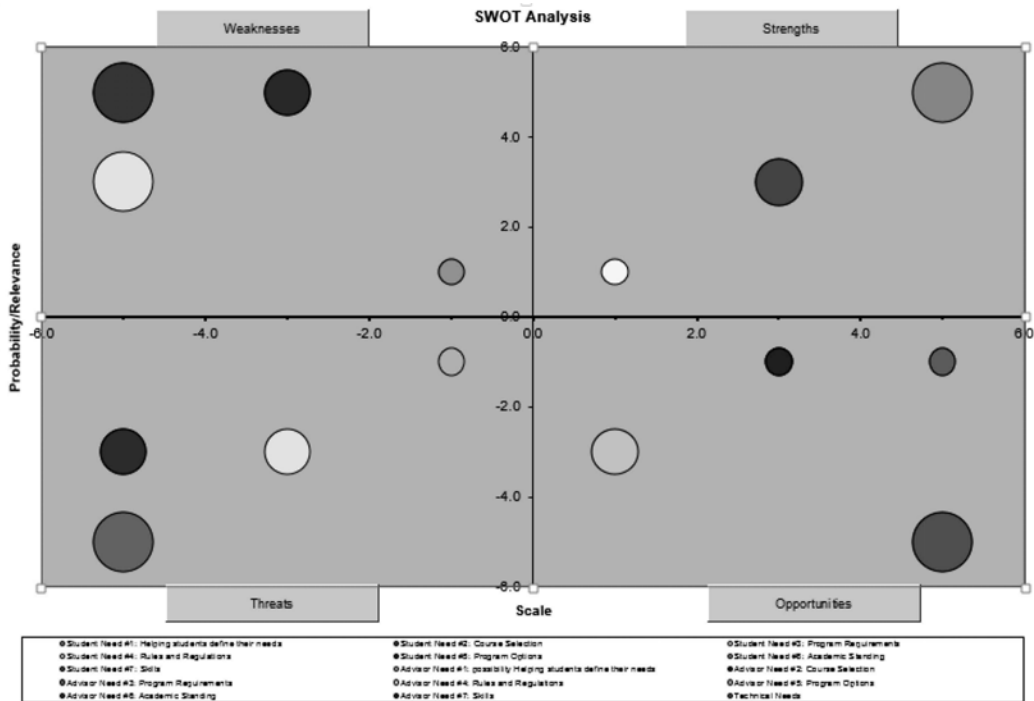
Z-axis

In our integrated application for KAU, SWOT analysis is used as a part of the process of the strategic decision. That context it will fit properly to consider the impact of the strategic decision the SWOT analysis factors on the system. The mapped impact shows the follows:

- The impact: High 5
- The impact: Medium 3
- The impact: Low 1

SWOT Chart

	Rating	Probability	Impact
Student Need #1: Helping students define their needs	5.0	5.0	5.0
Student Need #2: Course Selection	3.0	3.0	3.0
Student Need #3: Program Requirements is sufficient	1.0	1.0	1.0
Student Need #4: Clear Rules and Regulations	-5.0	3.0	5.0
Student Need #5: Program Options and varieties	-3.0	5.0	3.0
Student Need #6: Academic Standing	-1.0	1.0	1.0
Student Need #7: Skills to use the current system	5.0	-5.0	5.0
Advisor Need #1: Helping students define their needs	1.0	-3.0	3.0
Advisor Need #2: Course Selection support	3.0	-1.0	1.0
Advisor Need #3: Program Requirements	-5.0	-5.0	5.0
Advisor Need #4: Rules and Regulations	-3.0	-3.0	3.0
Advisor Need #5: Program Options	-1.0	-1.0	1.0
Advisor Need #6: Academic Standing	-5.0	5.0	5.0
Advisor Need #7: Staff Skills for the current system	-5.0	-3.0	3.0
Technical Needs	5.0	-1.0	1.0



Interpretation

Items mapped/plotted close to the (0,0) have least significance, since they are not mapped for high importance or the relevance or probability is very low.

Those that are at the extremes of the chart are rated as most important and are rated as highly relevant or a high probability of occurring.

The size of the bubble indicates the strategic impact of the SWOT factor.

6 RESEARCHES FUTURE DIRECTIONS OR RECOMMENDATIONS

The conclusions of this paper is interrelated. The integration of the Academic Advising with the KAU SIS on an online basis would definitely improve the efficiency in managing and maintaining student full information (academic as well as bio and administrative information), thus it is highly needed to a better student services for KAU. It is therefore highly recommended to give a full consideration for implementation in the near future. Of course that development can be done through the following recommendations:

1. The academic policy and bylaws of KAU to acknowledge the necessities of such improvement and integration.

2. A brainstorming should be conducted to all users/beneficiaries of the system to draw some more recommendations and improvement of the current KAU SIS.
3. Security and control measures should have the proper attention to secure and maintain the system in an its integrated model.
4. An integrated Online KAU SIS should be implemented in accordance with the real needs of the academic advising so that the perceived benefits of administration and academic personnel, and students would be realized.
5. An assessment would be done after the implementation of the system in order to have a proper test of integrating system requirement in line with the academic advising growing requirements.

7 CONCLUSION

The SWOT analysis of the paper, prove the following conclusions as drawn.

1. The existing KAU SIS almost work in a semi computerized model; means still using paper-based transaction processing system, that will not support the seamless integration with the academic advising functionalities model.

2. Problems encountered in the existing KAU SIS were inefficient, error-prone, and costly maintaining for each of student information.
3. Information requirements needed online by students or advisors are student's personal details, program of study, academic achievements, that is not available properly.
4. The benefits of an online KAU SIS would be efficiency and cost effectiveness in conjunction with the academic advising information system.

REFERENCES

- Andres, C. (2005). *Admission System and Online Examination for Benguet State University*. Undergraduate Project Study. Benguet State University.
- Butag, F. (2007). *Benguet State University Elementary Laboratory School Enrolment and Pupil Information System*. Undergraduate Project Study. Benguet State University.
- Capron, H., & Johnson, J. (2004). *Computers: Tools for Information Age*. Pearson Education Inc.
- Catherine, A. (2007). *Rapid Application Development: A Quick Methodology*. Retrieved from www.ezinearticles.com
- Coorough, C. & Shuman, J. (2008). *Multimedia for the Web: Revealed*. Thomson Asia PteLtd.
- Desousa, T. (2008). *Benefits of Web Applications*. Retrieved from www.articlesbase.com.
- Florencio et al., (2008). *Information and Communication Technology in Basic Education*. Anvil Publishing Inc.
- Kendall, K. And Kendall, J. (2005). *System Analysis and Design*. Sixth Edition. Pearson Education Pte Ltd.
- Kelly, R., & Turban, E. (2009). *Introduction to Information System*. Second Edition. John Wiley & Sons (Asia) Pte Ltd. Pages 162–163.
- Kimmel, P. (2005). *UML DYMYSTIFIED* (pp 32). McGraw—Hill/Osborne.
- Lebajan, J. & Esporlas, E. (2008). *HTML and Web Design*. Jemma Inc.
- Marrero, S. (2007). *Student Information System for the University of the Cordilleras*. A master project study. University of the Cordilleras.
- Massachusetts Institute of Technology (MIT). (2009). *About WebSIS*. Retrieved from <http://student.mit.edu/cgi-docs/govwbin1.html>.
- Meloni, J. (2004). *PHP 5: Fast & Easy Web Development*. Jemma Inc.
- Schwalbe, K. (2008). *Information Technology Project Management*. Brown Madonna Press Inc. Sklar, J. (2006). *Principles of Web Design*. Thomson Learning, Inc.
- Swartz, N. (2007). *Data Management Problems Widespread: Organizations Should Regard Data as their Greatest Asset and Invest in Data Management Accordingly. (ON THE EDGE: The Use & Misuse of Information)*. Retrieved from www.accessmylibrary.com.
- Villafania, A. (2007). *CHED, NPO sign agreement to secure academic records*. Retrieved from Inquirer.ne.

Author index

- Abd El-Azeem, M.H. 377
Abdel Aziz Ali, N. 639
Abdessalem, W.B. 439, 447
AbdulAziz, A.B. 533
Abdullah AlSelouly, R.S. 321
Abdullah, M. 615, 623, 633,
639, 649, 659, 671, 773
Abdullah, M.A. 687
Abinader, R.A. 269
Abo-Rizka, M. 573
Abu Sharha, M.M. 491
Aknin, N. 543
Al Achhab, M. 151
Al Mosaar, M.S.A. 551
Al Mozaini, M.A.S. 299
Al Selami, F.A. 463, 471
Al-Afghani, S.A. 433
AlAmri, A. 649
Al-Amri, J. 557
Alaoui Harouni, H. 145
Al-Asmari, S. 615
Al-Azawi, R. 235
Albogami, H.M. 359, 365, 433
Algashmari, W.F. 633
Alghamdi, Y. 671
Alharbe, N. 623
Alharbi, I.M. 517
Al-Harithy, M.M. 313
Al-Johani, A.A. 527
Al-Khalidi Al-Maliki, S.Q. 417,
425
Alkhamash, E. 439, 447,
453
AlKhudairi, N.J. 339
Almajhaddi, A. 395, 405
Al-Msloum, A.S.H. 527
Almutairi, S. 395, 405
Al-Obaidy, M. 235
Alotaibi, N.M. 687
Alsaadi, A.A. 773
Alsadi, M. 681
Alshuaibi, B. 681
Alsulami, O.A.A. 527
Altamimi, A.B. 385
Altheyab, M.S. 633
Althobaiti, A.S. 659
Altuwairqi, K.A. 773
Alyateem, A.M.A. 327
Al-Yateem, A.M.A. 765
Alyoubi, A.A. 517
Alyoubi, B.A. 537
ALZahrani, E.H. 353, 483,
497
Amal, Y. 37, 45, 53
Amina, G. 139
Aoulad Abdelouarit, K. 543
Aqel, M.J. 287
Aref, M. 65
Ayesh, A. 235
Ba Shammakh, E.S. 773
Bagais, W. 681
Bahurmoz, N. 681
Baiocchi, O. 751
Bajdor, P. 703
Bardessi, H.G. 305
Bassiri, M. 29, 33
Basudan, A. 681
Bawazeer, K. 335
Bayahya, A.Y. 773
Belaouad, S. 37, 45, 53
Belkacem, K. 159
Bella, F. 203
Bella, S. 203
Ben Salamh, F. 257
Benaouda Chaht, A. 73
bin Hamid, N.B. 765
Boukelif, A. 139
Boulmane, L. 61
Bounoua, A. 73
Brzozowska, A. 711
Bubel, D. 717
Bui Thi, A.-H. 261
Butler, M. 453
Cakula, S. 565
Castelnuovo, G. 207
Castelnuovo, G. 215
Chekhlatyi, O. 731
Chmielarz, W. 87
Cirillo, F. 599
Costa, C.W.A. 167
Costa, J.C.W.A. 167
Cristea, C. 453
Do, D.-T. 261
Dou, W. 187
Dridi, K. 439
El Emary, I.M.M. 537
El Ghouch, N. 151
El Hissi, Y. 79
El Mohajir, B.E. 151
Elarif, T.I. 389
El-Awadi, R. 573
ElBagoury, B.M. 557
ElBahnasy, K. 757
Elbattah, M. 65
Elfahime, B. 61
ElGouz, H. 377
Elshahed, M.A. 389
En-Naimi, E.M. 151
Essa, E.I. 131
Fazio, P. 599
Foster, J. 295
Fouad, F. 477, 783
Fournier-Viger, P. 241
Furduescu, B.-A. 171
Gedeon, P.G. 269
Ghizlene, S. 159
Goher, M. 249
Grondys, K. 725
Hachem, E. 145
Haqiq, A. 79
Hashim, H. 295
Hong, T.-P. 241
Humaidan, S.S. 345
Ibrahim, A.E. 389
Kadhuk, M. 725
Kalinichenko, A. 731
Kasparova, E. 583
Khair, M. 107
Khalil, L.J. 107

- Khedr, A.M. 115
 Khedr, A.Y. 371
 Khedr, M.E. 195
 Kone, I. 61
 Kościelniak, H. 737
- Lattas, A.M.A. 633
 Lefebvre, O. 279
 Li, T. 241
 Lin, A. 295
 Lin, J.C.-W. 241
 Liu, X. 751
 Łobaziewicz, M. 697
 Lupia, A. 609
- Marano, S. 597
 Mauri, G. 203, 207, 215
 Mercuri, V. 203
 Mikulec, M. 221
 Miranda, A.M.L. 167
 Mohamed, D. 159
 Mohamed, I.I. 131
 Mohamed, R. 37, 45, 53
 Monem, A.A. 11
 Moussted, M. 29, 33
 Murgia, F. 203
 Muryjas, P. 97
 Mustapha, B. 37, 45, 53
- Nasser, A.A. 195
 Nguyen, H.-S. 261
 Noordin, N.H. 131
- Omar Khaldoun, A. 131
 Omar, D.M. 115
- Pleshkova, A.Yu. 179
- Radouani, M. 61
 Ramadan, R.A. 371, 377
 Ratniece, D. 565
 Razek, M.A. 305
 Rizka, M.A. 249
 Rocha, C.A.J. 167
 Rosenburg, D. 235
 Roushdy, M. 65, 557
 Rozhon, J. 221
- Saadi, H. 227
 Sabry, E.S. 377
 Sadek, F.M. 125
 Salem, A.-B.M. 65
 Santamaria, A.F. 597, 599
 Santoro, E. 207, 215
 Sbihi, B. 543
 Seghroucheni, Y.Z. 151
 Serianni, A. 599
 Shahab, S.N. 131
 Sicurello, F. 203, 207, 215
 Skoczyńska-Prokopowicz, B. 17
 Sobiczewska, B. 23
 Soler Costa, R. 1
 Soler Santaliestra, J.R. 1
 Starostka-Patyk, M. 737, 745
 Strangis, F. 589
- Szumski, O. 87
 Szymczyk, K. 711
- Tagliente, I. 203
 Taieb, M. 681
 Talbi, M. 29, 33
 Touhami, R. 227
 Tropea, M. 589, 597, 599
- Voznak, M. 221, 241, 261
- Wang, X. 187
 Wawer, M. 97
- Yagoub, M.C.E. 227
 Yalid, A.T.A. 29, 33
 Yousef, M.I. 497, 503, 511
- Zainun, A.R. 131
 Zaitoun, N.M. 287
 Zaky, M.H. 195
 Zawada, M. 745
 Zhang, L. 187
 Zhukova, K.V. 179
 Ziti, C. 145
 Zoppis, I. 203, 207, 215
 Zouaoui, C. 73