Lorna Uden · Marjan Heričko
I-Hsien Ting (Eds.)

LNBIP 224

Knowledge Management in Organizations

10th International Conference, KMO 2015
Maribor, Slovenia, August 24–28, 2015
Proceedings

Springer

Lecture Notes in Business Information Processing 224

Series Editors

Wil van der Aalst
Eindhoven Technical University, Eindhoven, The Netherlands

John Mylopoulos
University of Trento, Povo, Italy

Michael Rosemann
Queensland University of Technology, Brisbane, QLD, Australia

Michael J. Shaw
University of Illinois, Urbana-Champaign, IL, USA

Clemens Szyperski
Microsoft Research, Redmond, WA, USA

More information about this series at http://www.springer.com/series/7911

Lorna Uden · Marjan Heričko
I-Hsien Ting (Eds.)

Knowledge Management in Organizations

10th International Conference, KMO 2015
Maribor, Slovenia, August 24–28, 2015
Proceedings

Springer

Editors
Lorna Uden
Staffordshire University
Stoke-on-Trent
UK

I-Hsien Ting
National University of Kaohsiung
Kaohsiung City
Taiwan

Marjan Heričko
University of Maribor
Maribor
Slovenia

ISSN 1865-1348 ISSN 1865-1356 (electronic)
Lecture Notes in Business Information Processing
ISBN 978-3-319-21008-7 ISBN 978-3-319-21009-4 (eBook)
DOI 10.1007/978-3-319-21009-4

Library of Congress Control Number: 2015943796

Springer Cham Heidelberg New York Dordrecht London
© Springer International Publishing Switzerland 2015
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media
(www.springer.com)

Preface

The 10th Knowledge Management in Organizations (KMO) Conference: Knowledge Management and Internet of Things was held during August 24–28, 2015, at the University of Maribor, Slovenia.

The Internet of Things (IoT) involves using sensors and actuators to track and manage machinery and other physical assets across a network. It is rapidly gaining momentum, bringing millions of devices and objects into the connected world and enabling whole new ways of managing assets and operations. The IoT also has the potential to transform manufacturing, build infrastructure, provide health care, and manage supply chains by monitoring and optimizing activities and assets at a very granular level. The explosion of IoT applications provides many opportunities for businesses to improve performance and create new service offerings. The unparalleled connectivity among devices offers new convenience for users and consumers.

The proliferation of networked sensors through the IoT is creating more data. The large amounts of data collected, in turn, create new business opportunities for enterprises including OEMs, service providers, and software publishers, for the monetization of IoT.

To match this growth in data requires advanced analytics. There are many emerging challenges for creating the IoT. These include the integration and management of heterogeneous data, the integration and transfer of enriched data, the effective use of knowledge-based decision systems, retrieval and sharing of knowledge automatically from huge volumes of data, ensuring security and protecting privacy.

The data created from the IoT are only beneficial to organizations when they can be turned into useful knowledge. When data are turned into knowledge, the enterprise is better positioned to respond and innovate in all phases of its operation so as to gain competitive advantages and even build entirely new business models.

This growing capability of firms to derive meaning using data from the IoT means that knowledge management (KM) systems can provide the platform for companies to improve their most valuable asset. KM systems should form a crucial part of IoT investment. KM has the ability to integrate and leverage information from multiple perspectives. The IoT is uniquely positioned to take advantage of KM processes and procedures. These processes and procedures enable IoT applications to provide a rich structure so as to enable decisions to be made in a multitude of ways. Organizations do not make decisions based on one factor; the total picture is what should drive decisions. KM enables organizations to take the total picture IoT offers, and along with leveraging tools that provide processing speed to break up the data into subsets for analysis.

Developing a superior capacity to take advantage of the IoT will enhance competitive advantage through KM that will lead to improved services. KM systems should form a crucial part of IoT investment because it has the ability to process the type of knowledge that data from IoT can transform and exploit. Turning data from IoT into useful knowledge for real-time IoT analytics poses many new challenges to the development of IoT applications.

The KMO conference series provides a forum for scientists and practitioners from around the world who are active in the knowledge management to share knowledge. This was the tenth anniversary of the conference. The KMO conference has being growing steadily since its inception. The first conference originated from the knowledge sharing and collaboration between researchers from four universities, Staffordshire University in the UK, the University of Vaasa in Finland, Kamurai Madurai University in India, and the University of Maribor in Slovenia. The first conference was held at the University of Maribor, Slovenia. Since then, KMO conferences have been held in several different countries. The second KMO was organized in Italy by the University of Lecce; the third KMO in Finland in 2008 by the University of Vaasa and University of Applied Sciences in Vaasa; the fourth KMO in 2009 in Taipei, Taiwan; the University of Veszprem hosted the fifth KMO 2010 conference in Hungary, while the sixth KMO in 2011 was hosted by the Tokyo Institute of Technology. During the seventh conference in Salamanca, the first LTEC workshop was held alongside KMO; LTEC has being a part of KMO since. In 2013 the eighth KMO was held in Taiwan for the second time, hosted by the National University in Kaohsiung. The ninth KMO in 2014 and the third LTEC workshop were organized in Santiago de Chile. This year the conference returned to the place where it began in Slovenia.

The KMO conference is unique in several ways. Firstly, the conference aims to provide a holistic view on KM from different perspectives, including organizational, social, and technical, as well as from a business and economic perspective. Secondly, the conference includes different aspects of KM from diverse disciplines. Thirdly, the conference combines concepts, approaches, methods, and theories from different domains and disciplines, whereas pilot projects and case studies assist in identifying and establishing best practices. Fourthly, the KMO conference also brings state-of-the-art research into KM. Last but not least, KMO is a very friendly conference that provides an environment where participants can share and exchange knowledge as well as collaborate in research.

As in previous years, the papers from this year's proceedings address KM challenges and best practices related to KM in organizations. We had 59 papers this year. All the papers published in the proceedings have undergone a rigorous review process involving at least three reviewers. Authors of these papers come from 27 different countries, including Australia, Austria, Brazil, China, Colombia, Czech Republic, Finland, Germany, Hong Kong, India, Italy, Japan, Kenya, Korea, Malaysia, Oman, Serbia, Singapore, Slovakia, Slovenia, South Africa, Spain, Taiwan, UAE, UK, and USA.

The papers are organized into nine thematic sections as:

- KM Processes
- Successful Knowledge Sharing and KM Practices
- Innovations for Competitiveness
- KM Platforms and Tools
- Social Networks
- Intelligent Systems

- KM and IoT
- Knowledge Management in Health Care
- KM in Education and Research

We would like to thank our authors, reviewers, and Program Committee for their contributions and the University of Maribor, Slovenia, for hosting the conference. Special thanks to the authors and participants at the conference. Without their efforts, there would be no conference or proceedings. We would also like to express our thanks to Saša Kuhar, University of Maribor, for her work as local chair.

We hope that these proceedings will be beneficial for your research and that the information in this volume will be useful for further advancements the field of KM in both research and industry.

August 2015

Lorna Uden
Marjan Heričko
I-Hsien Ting

Organization

Conference Chair

Lorna Uden	Staffordshire University, UK

Program Chairs

Marjan Heričko	University of Maribor, Slovenia
I-Hsien Ting	National University of Kaohsiung, Taiwan

Local Chair

Saša Kuhar	University of Maribor, Slovenia

Program Committee

Akira Kamoshida	Yokohama City University, Japan
Albena Antonova	Sofia University, Bulgaria
Costas Vassilakis	University of the Peloponnese, Greece
Dai Senoo	Tokyo Institute of Technology, Japan
Dario Liberona	Universidad Santa Maria, Chile
Eric Kin-Wai Lau	City University, Hong Kong, SAR China
Fan Yang	Academy of Equipment, China
G.R. Gangadharan	IDRBT, Hyderabad, India
George Karabatis	University of Maryland, Baltimore County, USA
Guandong Xu	University of Technology Sydney, Australia
Flavia Santoro	Federal University of the State of Rio de Janeiro, Brazil
Hong Qing Yu	University of Bedfordshire, UK
Jari Kaivo-Oja	FFRC, Turku School of Economics, Finland
Li Weigang	University of Brasilia, Brasilia, Brazil
Luka Pavlič	University of Maribor, Slovenia
Marja Naaranoja	Vaasa University of Applied Sciences, Finland
Marko Bajec	University of Ljubljana, Slovenia
Matjaž Gams	Jožef Stefan Institute, Slovenia
Paul Horng-Jyh Wu	SIM University, Singapore
Reinhard Bernsteiner	Management Center Innsbruck, Austria
Remy Magnier-Watanabe	University of Tsukuba, Tokyo, Japan
Richard Self	University of Derby, UK
Rubén González Crespo	Universidad Internacional de La Rioja, Spain
Shahrokh Nikou	Åbo Akademi University, Finland

Victor Hugo Medina Garcia	Universidad Distrital Francisco José de Caldas, Colombia
William Wang	Auckland University of Technology, New Zealand
Wu He	Old Dominion University, USA

Local Committee

Lili Nemec Zlatolas	University of Maribor, Slovenia
Maja Pušnik	University of Maribor, Slovenia

Contents

Innovations for Competitiveness

Knowledge Management Platforms and Tools

Social Networks and Mining Techniques

Knowledge Management and the Internet of Things

Knowledge Management Processes

Knowledge Management in Organizations - A Bibliometric Analysis of Research Trends

Peter Kokol, Bojan Žlahtič, Grega Žlahtič, Milan Zorman, and Vili Podgorelec(✉)

Faculty of Electrical Engineering and Computer Science, University of Maribor, Smetanova 17, 2000 Maribor, Slovenia
{Peter.Kokol,Bojan.Zlahtic,Grega.Zlahtic, Milan.Zorman,Vili.Podgorelec}@um.si

Abstract. Knowledge management became an important part of our economy far before than the actual term "knowledge management". Many studies have been written about different aspects and viewpoints of knowledge management, however in this study we used a bibliometric mapping approach to determine the main research topics and the contexts in which they are employed. 10,599 information sources were retrieved from the Scopus bibliographical database using the search string "knowledge management" AND organization. Most information sources were published as conference papers or journal articles. The most productive period was from 2007 till 2011. United States, United Kingdom and China were the most productive countries. Four main research topics were identified: Education and healthcare, Techniques and systems, Knowledge management and Knowledge sharing. Chronologically, five periods emerged, namely: Infrastructural and organizational research, Technical issues and knowledge management systems in organizations, Enabling technologies, Practical applications of knowledge management in organizations, and Evaluation and validation of knowledge management practices.

Keywords: Knowledge management · Knowledge management in organizations · Research trends · Bibliometrics · Bibliometric mapping analysis

1 Introduction

Knowledge management (KM) practices have been an important part of our economy far longer than the actual term "knowledge management" started to receive considerable attention from both academic and economic sphere in late 1980s [1]. Different approaches have been invented throughout history which allowed people to pass knowledge to descendants or to share it in order to build on earlier experience. Modern concepts and practices of knowledge management, however, evolved throughout the last two decades, in times when the knowledge was recognized as a critical resource, vital for economic growth.

KM consists of the systematic processes for acquiring, organizing, sustaining, applying, sharing, and renewing all forms of knowledge, to enhance the organizational performance and create value [2]. KM is about acting to build and leverage knowledge

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 3–14, 2015.
DOI: 10.1007/978-3-319-21009-4_1

through an understanding of how it is created, acquired, processed, distributed, used, harnessed, controlled, etc. [3]. Therefore, knowledge management aims to facilitate the access, use, and reuse of valuable knowledge resources [4]. Effective KM involves learning to manage knowledge as both an object and as a process [2, 5], which requires executives to develop a general understanding of what knowledge is, as well as efficient and systematic methods for managing it within the organization [3].

Since the 1990s there has been an obvious shift from an information-based economy to a knowledge-based economy [1]. The success of business operating in an increasingly competitive marketplace of the knowledge-based economy depends critically on the quality of knowledge which those organizations apply to their key business processes [6]. Therefore, creation, management, and sharing of knowledge within the organization have become one of the important factors of the competitiveness that should not be overlooked by companies and organizations.

Simultaneously with the expansion of KM practices in organizations, a huge amount of research theories, topics and results have been published within the interdisciplinary field of knowledge management. KM publications in general focus on knowledge in organizations, knowledge-based, theory of the firm, strategy, and knowledge creation [7]. Even though KM discipline is relatively a new research discipline, it has already boasted a number of scientometrics research with the purpose of better understanding its identity. In this manner, [8] looked at the breadth and depth of the field, and searched for interdisciplinary connections among researchers. Chauvel and Despres [9] examined KM research area in six dimensions: phenomena, action, level, knowledge, technology and outcome. In [10] a meta-analysis has been applied to publications in three major KM journals (Journal of Knowledge Management, Journal of Intellectual Capital, and Knowledge and Process Management). In [11] authors extended this work by examining the most influential KM publications, and explored the specific issues of subjectivity and objectivity.

Dwivedi et al. [12] found organizational and systems context-based KM research are the most widely published topics. Chen and Xie [13] built an intellectual structure by examining a total of 10,974 publications in the KM field from 1995 to 2010. Document co-citation analysis, pathfinder network and strategic diagram techniques were applied to provide a dynamic view of the evolution of knowledge management research trends.

In this paper, however, we aim to provide an overview of the knowledge management in organizations (KMO) field using a bibliometric mapping approach of the KM in organization literature production. Bibliometric mapping approach is based on the text analysis of abstracts and relies on computer algorithms and visualization techniques [14]. Its results are term maps, in which terms are located in such a way that the distance between any two terms reflects the relatedness of the terms; terms are also grouped in clusters based on calculated term relevance scores [15]. In this way, a scientific landscape of main concepts, topics and terms in the KMO field will be created. Moreover, important relations between KMO topics and terms will be studied and identified. Interpretation of the maps will be based both on our experiences in the KMO field and the published literature. In this way, past, current and future research trends within KMO will be discussed.

2 Bibliometric Analysis

Bibliometrics could be described as an answer to the saying that ‘if you can’t measure it, you can’t manage it”. It became prominent because of the need to measure the effects of the large investments going into the research and development. Bibliometrics has its origins as early as the beginning of the last century, but, it became data-driven in 1964 with the introduction of the science citation index. Bibliometrics [16–18] analyses the properties of literature production in terms of measures like number of articles on a particular topic, the dynamics of literature production, most prolific source titles, most productive countries, institutions and authors and most cited papers. It could be formally defined as “the quantitative analysis of the bibliographic features of a body of literature” [19]. The idea is based on the assumption that most scientific discoveries and research results are eventually published in international scientific journals where they can be read and cited by other researchers. It uses quantitative methods for analyzing written documents. Bibliometrical studies are also used to examine the history and structure of a field, the flow of information into a field, the growth of the literature, patterns of collaboration amongst scientists, impacts of journals, and the long-term citation impact of a work [17].

A recent technique used in bibliometric analysis is bibliometric mapping [20] which visualize literature production based on word co–occurrences [7], co–citations, co–references, etc. A popular bibliometric mapping software is the VOSviewer[1] (Leiden University, The Netherlands) [21]. It creates so called term maps which express terms relatedness, associations between terms, and importance of terms.

2.1 Data Source and Corpus

Scopus (Elsevier, Netherlands) was selected as a bibliographical database from which the corpus was formed on February, 23rd, 2015. The search keyword string used was “knowledge management” AND “organization”. Search was performed in information source titles, abstracts, and keywords. All types of information sources written in English in the period 1977–2014 were included in the corpus.

2.2 Data Extraction and Analysis

Most productive countries, institutions, source titles, literature production dynamics and research subjects were extracted by Scopus built-in functions. Scopus records including information source abstract year of publication were pre-processed in Excel (Microsoft, USA) and exported to VOSviewer (Leiden University, The Netherlands) for bibliometrics mapping analysis.

[1] VOSviewer – Visualizing scientific landscapes, http://www.vosviewer.com/Home.

3 Results and Discussion

The corpus consisted of 10,599 information sources. As shown in Table 1, most of the information sources were presented as conference papers followed by journal articles. The large number of conference papers shows that the knowledge management in organisations is still forming its body of core literature on one hand and on the other hand that it is in rapid development as a scientific field needing rapid publication of results at conferences and critical and fast appraisal and validation of ideas on discussions during conference paper presentations.

Table 1. Types of information sources

Type of information source	Number of information sources
Conference paper	5,111
Article	4,337
Review	496
Book chapter	352
Conference review	114
Book	86
Article in press	30
Short survey	30
Editorial	23
Note	16

The dynamics of literature production presented in Fig. 1 shows a positive trend in the period 1996–2010, then a rapid decline in the number of published papers. On the first sight this observation might seem in contradiction with the statement above that the filed is still in development, and thus the number of information sources should be increasing. Figure 2 shows the literature production dynamics separately for articles and conference papers, and while the publication dynamics of conference papers shows a strong negative trend, the dynamics of article production is more stable and even increasing in the year 2013.

As expected most productive source titles (Table 2) are categorized as both conference proceeding and journals from computer science, knowledge management and artificial intelligence fields. The articles are not yet published in top journals (mostly journals are ranked in the second or third quarter of all journals), which confirms our thesis that the knowledge management in organisation field is still developing its core research literature production.

Not surprisingly most productive countries (Table 3) are also the most developed, and industrialized ones. Top ten countries represent the 65.1 % of all research literature production. These results reveal that countries that have shown an immense interest in the KMO research area are generally the same countries that have a healthy and competitive economy.

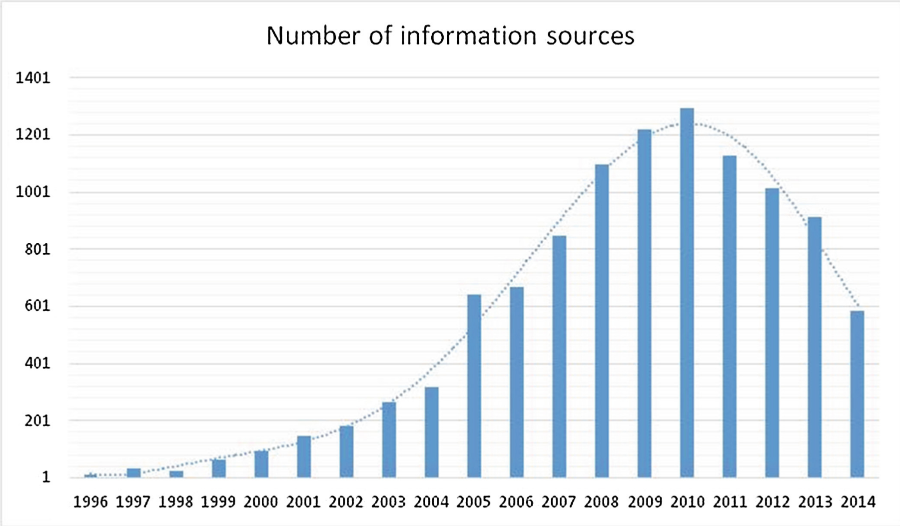


Fig. 1. The dynamics of research literature production for all information sources

The most productive institutions (Table 4) are coming from the most productive countries, with the exception being Daneshgahe Azad Eslami, which is located in Iran. Among top ten most productive institutions there are six from South East Asia: one from Singapore (National University of Singapore), two from Hong Kong (Hong Kong Polytechnic University and City University of Hong Kong) and even three from

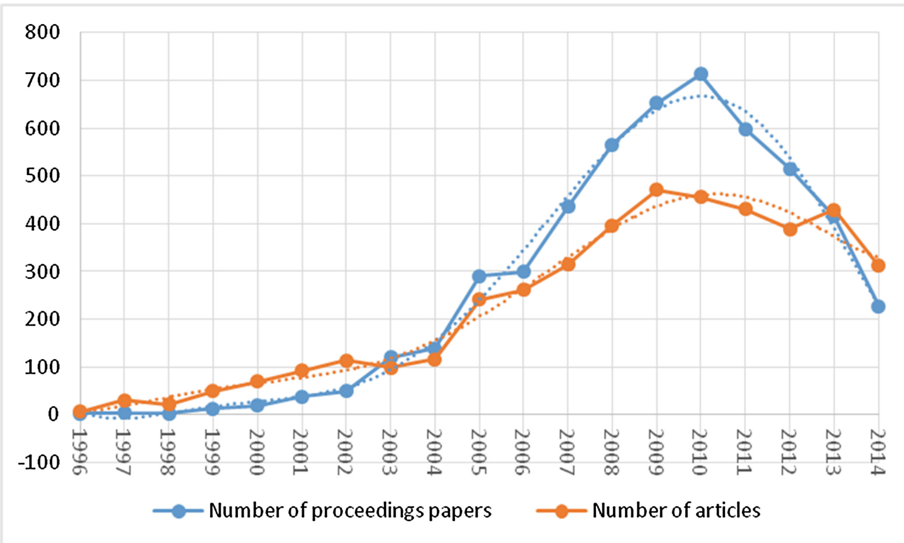


Fig. 2. The dynamics of research literature production for conference papers and articles

Table 2. Ten most prolific source titles

Source title	Number of information sources	SciMago journal rank
Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)	293	Proceedings
Journal of Knowledge Management	280	13,989
Vine	109	9,150
Journal of Information and Knowledge Management	106	18,564
Communications in Computer and Information Science	92	19,048
IFIP Advances in Information and Communication Technology	87	15,988
International Conference on Information and Knowledge Management Proceedings	84	Proceedings
Knowledge Management Research and Practice	77	9,163
Expert Systems with Applications	71	2,313
Learning Organization	68	9,364

Malaysia (Multimedia University, Universiti Teknologi Malaysia, and Universiti Teknologi MARA), which may be a bit surprising. It is also interesting that there is only one institution among top ten which is from USA (George Washington University), even though the USA is (by far) the most productive country within the field of KMO research.

Table 3. Ten most productive countries

Country	Number of information sources
United States	2,091
United Kingdom	1,135
China	802
Australia	608
Germany	491
Malaysia	379
Canada	370
Taiwan	355
Spain	341
Italy	336

The research in knowledge management in organisations is mostly focused on computer science, business, management, accounting, engineering, decision sciences and social science and also a bit surprisingly with health related research subjects (Table 5). It is interesting to see, that the number of information sources are almost

Table 4. Ten most productive institutions

Institution	Number of information sources
Loughborough University	89
Daneshgahe Azad Eslami	75
National University of Singapore	72
Hong Kong Polytechnic University	65
Multimedia University	65
Universiti Teknologi Malaysia	61
City University of Hong Kong	59
Universiti Teknologi MARA	50
University of Salford	48
George Washington University	46

Table 5. Ten most productive research subjects

Research subject	Number of information sources
Computer Science	4537
Business, Management and Accounting	3287
Engineering	2272
Decision Sciences	2070
Social Sciences	1996
Mathematics	578
Medicine	419
Economics, Econometrics and Finance	258
Biochemistry, Genetics and Molecular Biology	181
Environmental Science	178

evenly distributed between social sciences, decision sciences, and engineering, which confirms the interdisciplinary manner of knowledge management in organizations. Business, management and accounting as a driving force and computer science as an enabling field reign at the top of the list.

Bibliometric mapping analysis created four clusters (Fig. 3). Based on the terms we assigned a research topic to each cluster and if necessary divided each topic to sub-topics. In this way, four topics were defined:

- Education and healthcare (yellow cluster)
  - Healthcare (medicine, health, healthcare organization,. nurse, patient, hospital, evidence)
  - Education (student, faculty, education, center)
- Techniques and systems (red color)

– Web (web, access, digital library, e-learning, security, semantic web, information need, ontology, metadata, wiki, social network)
– Document management (document, text, graph, annotation, keyword, language, data set, domain, reuse, artefact)
– Management system (project manager, risk management, task, business rule, technique, project management, knowledge mapping, business process, supply chain, business goal)
– Software (software engineering, software development, software development process, iso, process improvement)

• Knowledge management (blue colour)
  – Activities (knowledge management process, knowledge creation, km process, km activity, knowledge activity, knowledge conversion, socialisation)
  – Entities (organisational culture, explicit knowledge, medium sized enterprise, competitive intelligence)
• Knowledge sharing (green colour)
  – Customer (customer, customer knowledge, crm – customer relation management, alliance, marketing, knowledge integration)
  – Resources (dynamic capability, km capability, organisational capability, social capital, organisational performance, generalizability)

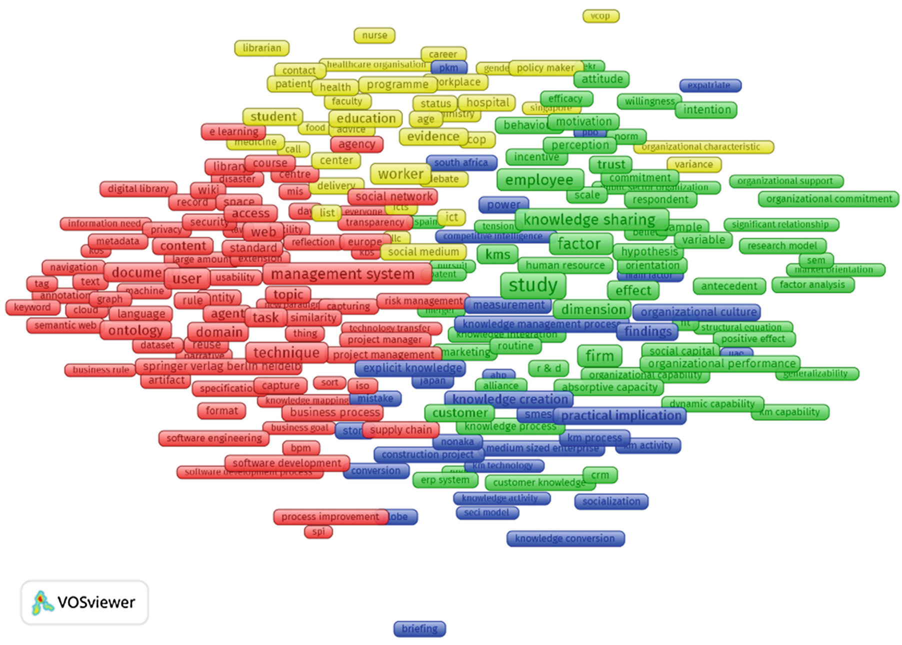


Fig. 3. The term map of all terms with more than 15 occurrences. Four clusters, representing main research topics, were identified: Education and healthcare (yellow), Techniques and systems (red), Knowledge management (blue) and Knowledge sharing (green) (Color figure online).

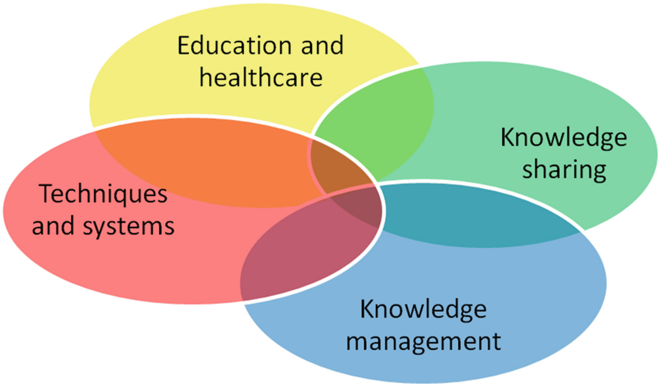


Fig. 4. The four identified clusters of knowledge management in organizations research.

– Motivational factors (trust, commitment, organisational support, motivation, willingness, attitude, efficacy, norm, behaviour)
– Models (factor, variable, dimension, effect)

The four identified clusters of KMO research are depicted on Fig. 4.

The results of bibliometric mapping analysis of KMO from the chronological point of view are presented on Fig. 5. The analysis was focused on the most productive

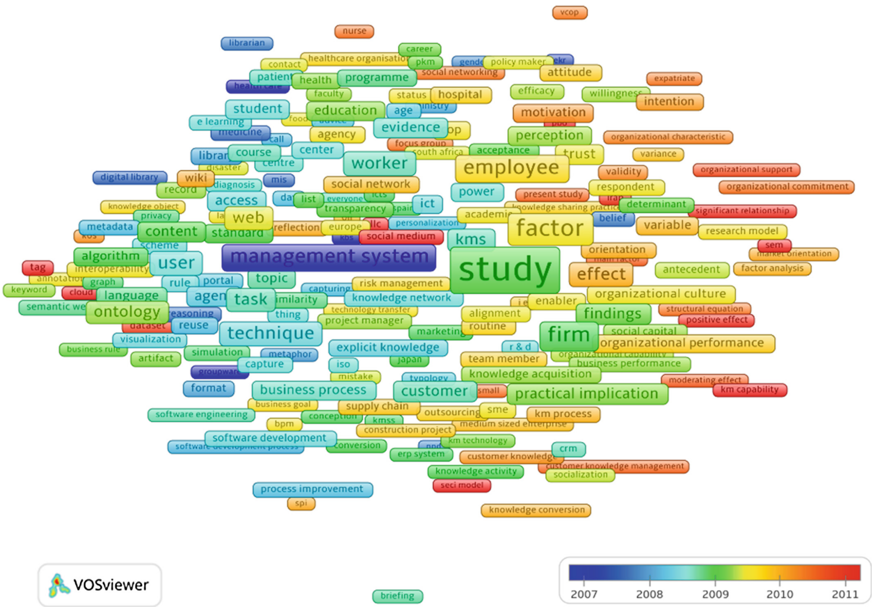


Fig. 5. The term map of all terms with more than 15 occurrences throughout time based on publication date. Five periods emerged: infrastructural and organizational research, technical issues and KM systems in organizations, enabling technologies, practical applications of KM in organizations, and evaluation and validation of KM practices.

period between 2007 and 2011 (see Fig. 2). In this way 5 periods of KMO research were identified.

It can be seen that in the earlier period of KMO research mainly infrastructural and organizational topics were addressed (management systems, groupware, library, digital libraries). It is interesting that in that early period KM research was primarily oriented towards medical and healthcare organizations (healthcare, medicine, and medical information systems were, and still are for that matter, important research terms of KMO research).

Then, these fundamental questions were expanded and deepened mainly in two directions: some researchers focused on technical issues on how to put KM into practice (the main terms being techniques, software development, software engineering, reuse, visualization, capture, access, metadata, explicit knowledge), while the others focused on users of KM systems in organizations (the main terms being user, customer, worker, student, patient, personalization).

Having the infrastructure, developed methodologies, enabling technologies, identified processes and educated users, the research trends then focused on practical applications of KM in organizations and setting the goals toward making use of it. In this manner, publication topics focused on general aspects (employees, team members, organizational culture, organizational performance, business goals, social networks, trust) as well as specific fields (hospitals, agencies, healthcare organizations, academia, medium sized enterprises, ERP systems) and/or regions (Europe, Japan, South Africa, Iran).

Finally, the most recent publications focus on evaluation and validation of KM practices, methods and technologies (the most frequent terms being variables, factors, factor analysis, significant relationships, KM capability, structural equation, validity, positive effects, moderating effects).

It can be seen that chronologically the following five periods emerged: Infrastructural and organizational research, Technical issues and knowledge management systems in organizations, Enabling technologies, Practical applications of knowledge management in organizations, and Evaluation and validation of knowledge management practices. They are depicted on Fig. 6.

Based on the evolution cycle presented in Fig. 6 we might expect that evaluation and validation of KM practices will catalyse new KM approaches, theories and practices leading to the beginning of a new evolution cycle. Some of the emerging technologies with viable perspective are media – supported knowledge management [23], semantically enhanced KM systems [24] and global knowledge management [25].

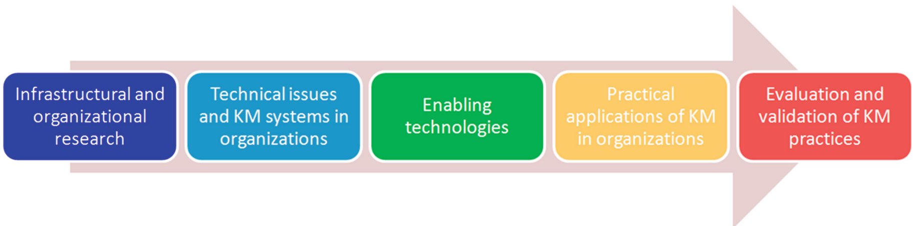


Fig. 6. The five emerged periods of knowledge management in organizations research.

4 Conclusions

In this paper we focused on knowledge management in organizations as a term in scientific publications, where we tried to identify the past and current state of this research area and its potential for the future. Not only have we highlighted all the different states of this research area, we have also identified all the main fields that KMO is applied to or has been researched in.

The results of the performed bibliometric mapping analysis reveal that countries that have shown an immense interest in this research area are the same countries, at least for the main part, that have a healthy economy. The results also showed that knowledge management has been, right from the beginning of more intensive KMO research, very much oriented toward medical and healthcare organizations. And it still is – according to [22], biotech industry firms are (by far) the most aware of the importance of knowledge management practices among all industries.

What is especially interesting and shouldn't be overlooked, are our findings that the whole research area has been undergoing a healthy evolutionary cycle. This is showing that KMO is evolving and maturing, which best explains the wide spread use and research of this area. Since we have drown quite some links, not only between the main areas that KMO is applied in, but also between different sub areas or sub domains, we can see different trends emerging and can even more clearly see their roots. Though we have detected a recent decrease in certain types of publications when it comes to KMO, we have observed that this is an indicator of maturing of a research field and should be understood as a sign that we can expect a probable drop in quantity and an increase in quality of further researches.

Acknowledgement. This work was produced in part within the framework of the operation entitled "Centre of Open innovation and ResEarch UM". The operation is co-funded by the European Regional Development Fund and conducted within the framework of the Operational Programme for Strengthening Regional Development Potentials for the period 2007–2013, development priority 1: "Competitiveness of companies and research excellence", priority axis 1.1: "Encouraging competitive potential of enterprises and research excellence".

References

1. Jing, F., Chakpitak, N., Goldsmith, P.: A knowledge supply chain: reengineering e-tourism curriculum design. e-Rev. Tourism Res. **10**(2), 56–62 (2012)
2. Choo, C.W.: The Knowing Organization: How Organizations use Information to Construct Meaning, Create Knowledge, and Make Decisions, 2nd edn. Oxford University Press, New York (2006)
3. Evans, M.M., Dalkir, K., Bidian, C.: A holistic view of the knowledge life cycle: the knowledge management cycle (KMC) model. Electron. J. Knowl. Manag. (EJKM) **12**(2), 85–97 (2014)
4. Dieng-Kunz, R., Matta, N. (eds.): Knowledge Management and Organizational Memories. Kluwer Academic Publishers, Boston (2002)

5. van den Berg, H.A.: Three shapes of organisational knowledge. J. Knowl. Manag. **17**(2), 159 (2013)
6. Macintosh, A., Filby, I., Kingston, J.: Knowledge management techniques: teaching and dissemination concepts. Int. J. Hum. Comput. Stud. **51**, 549–566 (1999)
7. Sedighi, M., Jalalimanesh, A.: Mapping research trends in the field of knowledge management. Malays. J. Libr. Inf. Sci. **19**(1), 71–85 (2014)
8. Ponzi, L.J.: The evolution & intellectual development of knowledge management. Ph.D. dissertation, Long Island University (2003)
9. Chauvel, D., Despres, C.: A review of survey research in knowledge management: 1997–2001. J. Knowl. Manag. **6**(3), 207–223 (2002)
10. Serenko, A., Bontis, N.: Meta-review of knowledge management and intellectual capital literature: citation impact and research productivity rankings. Knowl. Process Manag. **11**(3), 185–198 (2004)
11. Nonaka, I., Peltokorpi, V.: Objectivity and subjectivity in knowledge management: a review of 20 top articles. Knowl. Process Manag. **13**(2), 73–82 (2006)
12. Dwivedi, Y., Venkitachalam, K., Sharif, A., Al-Karaghouli, W., Weerakkody, V.: Research trends in knowledge management: analyzing the past and predicting the future. Inf. Syst. Manag. **28**(1), 43–56 (2011)
13. Chen, T.T., Xie, L.: Identifying critical focuses in research domains. In: Proceedings of 9th International Conference on the Information Visualization (IV 2005), pp. 135–142. London, 6–8 July 2005
14. Heersmink, R., van den Hoven, J., van Eck, N.J., van den Berg, J.: Bibliometric mapping of computer and information ethics. Ethics Inf. Technol. **13**, 241–249 (2011)
15. Waltman, L., Van Eck, N.J.: A new methodology for constructing a publication-level classification system of science. J. Am. Soc. Inf. Sci. Technol. **63**(12), 2378–2392 (2012)
16. De Bellis, N.: Bibliometrics and Citation Analysis. The Scarecrow Press, Inc., Lanham (2009)
17. Garfield, E.: The history and meaning of the journal impact factor. J. Am. Med. Assoc. **295** (1), 90–93 (2006)
18. Wenjuan, Z., Xuezhi, W., Tieqin, Z.: Progress analysis of library knowledge management research in China based on bibliometrics. J. Appl. Sci. **13**(9), 1484–1488 (2013)
19. Hawkins, D.T.: Bibliometrics of electronic journals in information science. Inf. Res. Int. Electron. J. **7**(1). http://informationr.net/ir/7-1/paper120.html (2001). Accessed 23 May 2013
20. Liu, D.D., Liu, S.L., Zhang, J.H.: Visualization analysis of research hotspots based on CiteSpace II: taking medical devices as an example. Med. Devices: Evid. Res. **7**, 357–361 (2014)
21. Van Eck, N.J., Waltman, L.: VOSviewer Manual Leiden. Univeristeit Leiden, Leiden (2013)
22. von Nordenflycht, A.: What is a professional service firm? Towards a theory and taxonomy of knowledge intensive firms. Acad. Manag. Rev. **35**(1), 155–174 (2010)
23. Zhang, X., Gao, Y., Yan, X., de Pablos, P.O., Sun, Y., Cao, X.: From e-learning to social-learning: mapping development of studies on social media-supported knowledge management. Comput. Hum. Behav. (2015). doi:10.1016/j.chb.2014.11.084
24. Cob, C., Abdullah, R., Risidi, H., Mohd, N.M.: Preliminary study on semantic knowledge management model for collaborative learning. ARPN J. Eng. Appl. Sci. **10**(2), 442–450 (2015)
25. Pawlowski, J., Pirkkalainen, H.: Global social knowledge management: the future of knowledge management across borders? In: 3th European Conference on Knowledge Management, ECKM 2012, pp. 933–942. Academic Conferences Limited, Cartagena (2012)

Unifying Knowledge Creation Process Through Cascading Modes of Communication

Paul Wu Horng-Jyh(✉)

School of Science and Technology, SIM University,
461 Clementi Road, Singapore 599491, Singapore
paulhjwu@unisim.edu.sg

Abstract. In this short paper, we elaborate the concept of Cascading Modes of Communication (CMC) for knowledge creation in social media. Through a review of CMC and Multi-Stakeholder Learning Dialogue (MSLD), the 3 distinct dimensions of knowledge creation processes: the Epistemological, the Ontological and the Communicative, are posited as reference for the analysis of the processes. Each SECI process is framed as a conversion of knowledge along all 3 dimensions. The relationship of Expansive Learning and SECI cycles is then explored. Two previously discussed SECI processes: Boom-up and Slip-down are explained as phenomenon observed in boundary crossing. Furthermore, two new SECI processes symmetric to Boom-up and Slip-down are identified: Boil-Up and Percolate-down, pointing to the pervading tension between tacit-explicit, local-global, and consensus-dissensus knowledge conversion processes. This tension provides a theoretical explanation for the self-sustaining nature underling all knowledge creation process. Thus, the social media scaffold in CMC is to function as boundary objects created to bridge the stages of knowledge creation processes.

1 Dimensions in SECI Communicative Discourses

In an earlier paper [1], we propose that SECI knowledge creation steps can be viewed as a communicative discourse with Cascading Modes of Communication (CMC) for knowledge creation in social media. Namely, viewed from a *Communicative Dimension*, each step corresponds to a particular Speech Act Pattern, with a distinct criteria for meaningfulness, and only appropriate in a certain kind of social media, as shown in Fig. 1.

As demonstrated in Mode 1 of Fig. 1, the main purpose of the speech acts for knowledge creation in Socialization is to convey the belief or emotion of an (or a group of) individual(s). For Socialization to be meaningful, the criteria is that the speech acts must be authentic, expressing the true feeling or belief of those involved in Socialization; *Epistemologically speaking*, this criteria is consistent with the tacit to tacit knowledge conversion as explained in the SECI framework. In Socialization, the participants can only take the "face value" of the actions meant by the other participants as each participant fully "accept" and "express" what's in others' and their own minds.

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 15–25, 2015.
DOI: 10.1007/978-3-319-21009-4_2

	SECI Space	Criteria of Meaningful-ness	Communication Mode (Speech Act Pattern)	Social Media and Reports
Mode 1	Socialization	Authenticity (truth and sincerity)	This, X, is truly what I believe/feel/like. (The expressives and constatives)	Blog or Social Network
Mode 2	Externalization	Rationality (rightness and satisfaction)	I agree/disagree with /need to clarify about X, because of Y. (The declaratives and imperatives, and interrogatives)	Forum
Mode 3	Combination	Normality (rightness and satisfaction)	Regardless of X, Z is what most of us believe/feel/like. (The directive-commissives, with concessions)	Wiki
Mode 4	Internalization	Reality	Given Z, how do I do about X. (This involves material actions rather than communicative actions.)	Reflective; Report

Fig. 1. Cascading modes of communication

That is, the force of Expressive and Constative expressed by the Sender (of the message) does not invite particular reaction on the part of the Interpreter (who receives the message)[1]. Here we can assume that the achievement of "consensus," namely, X, is taken for granted, and there is no misunderstanding in the communication among the participants. *Ontologically speaking*, Socialization typically occurs in the context of a small group. However, from a bird's eye view, though various small groups may have consensus within themselves, between different small groups, dissensus may remain. So when each small group come together, boundary crossing needs to happen before consensus can be achieved. Thus, a different Speech Act pattern needs to be adopted.

The main purpose of the speech acts in Externalization is to express agreement or disagreement with a particular participant's belief or feeling by giving justification or clarification as shown in Mode 2 of Fig. 1. Continuing from the above scenario where two small groups encounter each other, the consensus of the two groups cannot be taken for granted. When dissensus does occur, it requires clarification and

1 Speech Act is a schema involves several dimensions itself, including the "direction of it." A Speech Act may require the World (external and subjective) to be conformed to the Message, or the Message to the World. The forces associated with Expressives and Constatives indicate the "direction of fit" is from the Message to the World that requires no obligation from the interpreter to do anything with respect to the sender's Message. For further details please consult [2]. We also adopted the use of Sender and Interpreter from [2] to represent a Speaker and Listener involved in a communication, respectively.

argumentation to take place to resolve it. After each party examines the dissensus, if they think the dissensus cannot be resolved as such, they are required to give a justification or reason, namely, Y, why consensus cannot be achieved. If they think the dissensus can be resolved, they do the same. The justification is critical, and the discourse may progress requiring further background (Y_1, Y_2, etc.) to be foregrounded/clarified. It externalizes knowledge (explicit) hidden in the background (tacit) to allow further engagement so that the dissensus may be resolved. The justification also produces the force of a Directive[2], so the World (including the Interpreter) would conform to how the Sender would intend it to be. Nonetheless such force must be given in a logical way (or rationally), otherwise the Directive is not likely to be achieved. Viewed from an *Ontological dimension*, the scenario where Externalization occurs is not limited to between two small groups. In general, Externalization will be required as long as there is a boundary (social or historical) that needs to be crossed. For example, it could happen between an individual and another individual (due to different historical background). Or it could be between different levels of management in an organization (middle v.s. top managements). The main characteristic, viewed from an *Epistemological dimension*, is that in such discourse both parties use argumentation to foreground tacit knowledge that (meaningfully) may help to bridge the gap of mutual understanding to achieve common explicit knowledge. Now, for bigger grouping to happen along the Ontological dimension, another Speech Act pattern needs to be adopted.

As shown in Mode 3 of Fig. 1, at the Combination step, most relevant background knowledge (or tacit knowledge) pertaining to the point of discussion would have been foregrounded (made explicit), the focus is on what the explicit knowledge that has been commonly agreed is, without any prejudice. To demonstrate the extent to which bias should be avoided, as shown in the Speech Act pattern, a concession is given to say that even the Sender's own views are to be discounted – "Regardless of X," which is what the Sender truly felt or believed. It is only with this neutral point of view, that the common knowledge can be committed together (a Commissive Speech Act)[3]. *Epistemologically speaking*, the combination of explicit knowledge, resulting in Z, has been the focus of most of the modern information management systems. Technologies such as Big Data Analytics is just one of the most recent development where structured or unstructured data are gathered to mine for patterns in business or social trends based on statistical inferences. However, the basis of such combination must be based on sources of data without any bias, if any information or knowledge that is excluded due to prejudice then the validity of conclusion would be flawed. Thus the selection of what explicit knowledge to be combined must follow a norm that is accepted by the community at large where the groups are a part. Thus *Ontologically speaking*, the leadership of the community must first establish itself as a legitimate authority to make a statement for the community, representing the consensus of the community, and not

[2] The Directive's "direction of fit" is World to Message [2], that means, the World must conform to what is communicated in the Message. Since the Sender is part of the World, s/he is then acquired the obligation to make the World as it is intended by the Sender.

[3] A Commissive binds the Sender, together with its Interpreters, to a common course of action [3].

just for opinions of a particular interest group. Or put it another way, the one who is able to argue for and forge a commonly accepted knowledge will have the legitimacy to be the leader of the community/organization.

In Mode 4 of Fig. 1, when individuals are faced with the task of embodying Z in their individual practices, they may realize the discrepancy (which was X), if any. In such circumstance, the question: Given Z, how do I do about X, will naturally arise. If there is a great degree of consistency between Z and X, then the individual may find they can be in a leadership position to carry out what is intended by Z. Otherwise, they are obliged to be lead by the leadership in order to pursue Z. In these circumstances the Material Action, rather than Communicative Action, of the individual is what matters. The Material Action is targeted to bring about the desire conditions in the World that is consistent with Z. Ontologically speaking, an individual, or a small group, is being held accountable by a larger group or the entire organization against Z (which has been debated and accepted as a norm, at least for the time-being). Epistemologically speaking, the explicit knowledge is being converted to tacit knowledge as individuals, through developing new practices attempt to embody the new knowledge as the commonly accepted best practices.

As we complete the review of CMC, we have also realized that each step of a SECI knowledge creation process is not only performed with respect to the Epistemological dimension where the conversion between tacit and explicit knowledge takes place. It is also performed in an Ontological dimension where crossing layers of grouping: the individual, group, organization and social network levels, takes place. Most importantly, we also realize that in the Communicative dimension, each Speech Act is intended to bring about a consensus where dissensus exists. The situation is summarized in the following Table 1:

Table 1. Dimensions and scale in SECI knowledge creation process

Dimension	Scale	
Epistemological	Tacit	Explicit
Ontological	Local (smaller grouping)	Global (larger grouping)
Communicative	Dissensus	Consensus

Note that the proximity of tacit, local and dissensus knowledge is a by-product of human nature, as explicit, global and consensus is a result from efforts made through intentional communication that form part of the social realities that every human exists. In the next section, we shall review a theory, called Multi-Stakeholder Learning Dialogue (MSLD) that summarizes the same fact from various theories about organizational behaviors.

2 Sociological Research Paradigm and Multi-stakeholder Learning Dialogue

How individuals create knowledge in an organization can be compared to how sociological research is done in organizational sciences. In [4], Deetz posited two dimensions of contrast in the sociological research paradigm for organizations, which is based on the difference in discursive moves and social relations. The first dimension focuses on the origin of concepts and problem statements as part of the constitutive process in research. Local/emergent research conception is contrasted with "Elite/a priori". Similarly, we may say knowledge creation may be originated from an individual or from a group of individuals (the leader of the group, particularly), who have some privileged access to certain critical knowledge or insights. Research orientation can also be contrasted in the extent to which they work within a dominant set of structuring of knowledge, social relations, and identities, called a "consensus discourse," and the extent to which they work to disrupt these structuring, called the dissensus discourse. Once we have established the two dimensions, we arrive at the following landscape of representational practices in Sociological Research. As shown in Fig. 2, there are four distinct discursive discourses in the representation practice: Normative Studies, Interpretive Studies, Dialogic Studies, and Critical Studies. In [5], the author adapted the framework in Fig. 2 by suggesting that its four discursive discourse (or sensemaking) quadrants represent not only alternative ways of knowing but also potential stages in a conversational journey that enables theorists and practitioners to engage and comprehend the different social realities constructed by participants in a stakeholder network as shown in Fig. 3.

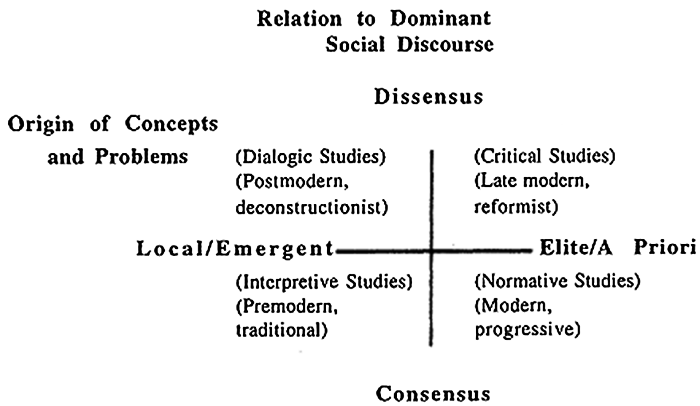


Fig. 2. Contrasting dimensions from the metatheory of representational practices

The authors then proceed to describe a Multi-Stakeholder Learning Dialogue (MSLD) that allow the managers to arrive a constructive ethics based on consulting the moral knowledge of all stakeholders in the network from moral persuasions and representational practices. They argued that "If we are to develop a coherent, credible

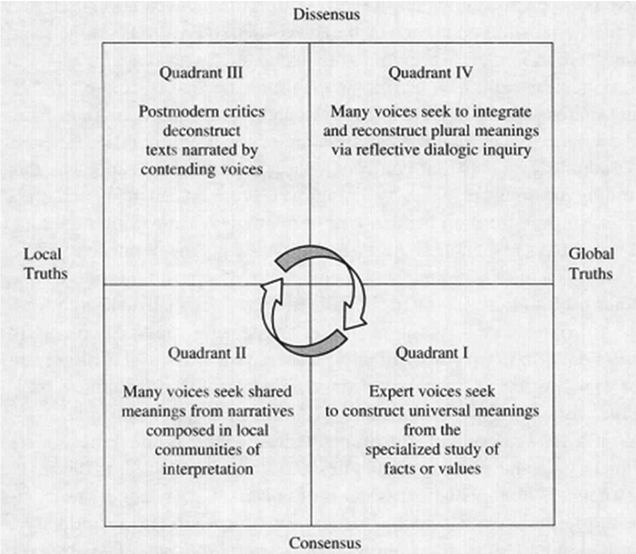


Fig. 3. Conversional journey to engage and comprehend paradoxical realities constructed by participants in a stakeholder network

representation of corporate citizen practice, we must get beyond Quadrant 1 efforts to dredge, sort and configure facts from the mixed record of pervious or existing organizational and social arrangements. Nor should we limit ourselves to deducing universal ethical values from hypothetical thought experiments (also Quadrant 1) in the hope that these value frameworks will shape the exercise of managerial discretion. We should not embrace without questioning the shared meanings constructed by local communities of interpretation in Quadrant 2. We should ask the tough questions about the paradoxical juxtapositions of observer and observed, facts and values and knowledge and power in the contested language games of Quadrant 3. In particular, we should be prepared to enter the transformational realm of reflective inquiry in Quadrant 4.

Here we can relate the spiral in Fig. 3 with the modes of in CMC if we align the Ontological dimension with the local truth and global truth dimension, and the Epistemological dimension with the dissensus (tacit) and consensus (explicit) dimension. Once the two dimensions are aligned, we arrive at the following 3 dimensional representations of Cascading Modes of Communication.

As shown in Fig. 4, it is noted that Quadrant I (Expert Voices seek to construct universal meaning) is mapped with Mode 3 (Combination/Commissive); Quadrant 2 (Many Voices seeking shared meaning), with Mode 2 (Internalization); Quadrant 3 (Postmodern critics deconstruct text), with Mode 1(Socialization/Expressives & Constatitives); Quadrant 4 (Multiple voices trying to integrate and reconstruct pluralistic meaning), with Mode 2 (Externalization/Directives). Figure 5 summarizes the mapping between SECI and MSLD.

These mappings may seem non-sensible at first sight. However, when we realize that the explanation given in MSLD and SECI actually assumes "opposite"

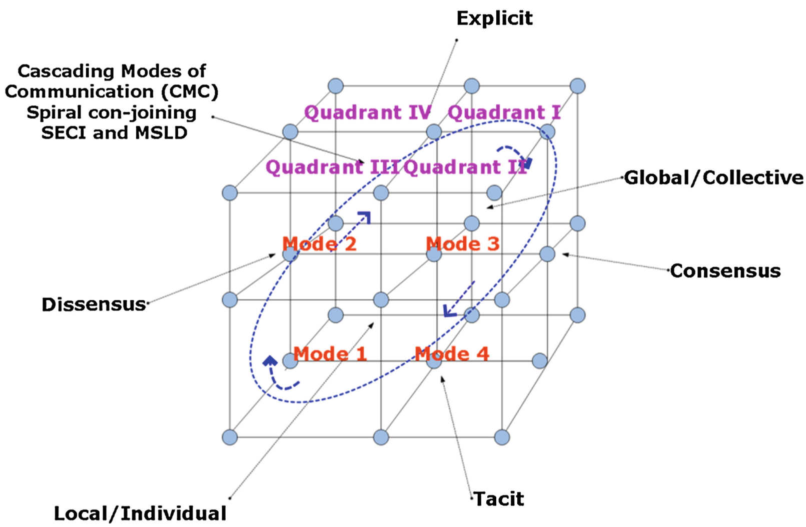


Fig. 4. Three dimensional CMC spiral joining SECI and MSLD

SECI	MSLD
Mode 1 (Socialization – Expressives & Constatives)	Quadrant 3 (Postmodern critics deconstruct text)
Mode 2 (Externalization – Directives)	Quadrant 4 (Multiple voices trying to integrate and reconstruct pluralistic meaning)
Mode 3 (Combination – Commisives)	Quadrant 1 (Expert voices seek to construct universal meaning)
Mode 4 (Internalization – Material Actions)	Quadrant 2 (Many voices seeking shared meaning)

Fig. 5. Mapping between SECI and MSLD

perspectives, the picture becomes clear. Namely, the theorizing of MSLD assumes the vintage point of the Organization (or the Management), while in SECI, it assumes that of the individuals (or the Employee). Once this is realized, we come to understand, for every knowledge creation step (or discursive discourse) there are always two, seemingly opposite, perspectives; namely, those of the Management's vs. those of the Employee's.

To be more specific, from a Management's point of view, a Combination (Mode 1) represents the most desirable outcome where every individual (employee) has to commit to a common knowledge. The designated Experts in the organization will then be able to expound the truth across the entire organization to achieve the objectives obtained from such combination of knowledge (Quadrant 1). When such truth is percolated down the organizational hierarchy, Internalization (Mode 4) needs to take place. However, Internalization not only requires the individuals embodying the corporate norms, reciprocally, it also means that the organizations needs to allow space for the accepted common knowledge (the truth) to be interpreted by the individuals to produce their respective, concrete meanings. Consequently, such contextualized meanings (tacit knowledge) in fact engender the creation of newer knowledge to complement (or contextualize) the accepted one. Internalization is followed by Socialization (Mode 1). In Mode 1, while the individuals are enjoying among their peers in their respective comfort zones, the Organization (or the Management) should be concerned whether such locally interpreted truths shared among individual employees are indeed aligned with the overall corporate goal intended by the common knowledge (the truth). It is likely that when the individualistic views (post-modernistic) deconstruct the text (the truth), which is regarded as the standard view, discrepancy will begin to emerge. Thus in Mode 2 (Externalization), it is when discrepancy of local truths becomes an issue that the individuals (the Employees) would be concerned enough to engage with each other in argumentation and debate (with the Management as well) in the hope that a common pluralistic meaning can be achieved (Quadrant 4). And such steps will repeat by itself until more and more tacit knowledge is surfaced, converted to explicit knowledge and combined, and then in turn internalized as tacit knowledge for the individuals and the process continues as the knowledge spiral continues to be traversed as needed.

3 Application of CMC to Re-analyzing SECI Case in View of Expansive Learning

In [6], an ontological shift SECI model was presented. The case on point was analyzed by two major kinds of SECI processes with reference to an ontological structure (namely, the hierarchy of an organization and social network): Booming-up and Slip-down. Booming-up is demonstrated to traverse upward in the ontological dimension that gathers momentum of a knowledge creation process. On the other hand, Slip-down is regarded as an opposite of Booming-up, where proposed new knowledge (from the lower organizational layer) was rejected at a higher layer. Thus, the introduction of the ontological dimension in SECI framework produces a distinction, namely, a *boundary* differentiating two groupings, and it is critical to the success of knowledge creation. However, concepts such as boundary and boundary crossing are foundational to Expansive Learning in the framework of Activity Theory. In what follows, we shall explore Expansive Learning to reveal its connection to SECI.

First, the expansive learning cycle (see Fig. 6) is used to clarify the state where the organization is at the moment (development challenge). Secondly, the cycle is used to guide the planning of the individual exercise sessions and discussions. The cycle consist of several learning acts: *questioning* present work practices – in CMC terms: expressing authentic but problematic feeling or belief resulting from problematical internalization of accepted truth, *analysing* historically the causes that have created problems in daily work – in CMC terms: clarifying and arguing a point by giving justification to doubts earlier expressed in socialization, *modelling* and *searching* for a new form of activity – in CMC terms: combing all argumentation to arrive an acceptable common ground to be committed together, *testing* and changing the activity and practices during the experimental phase and finally reflecting on the process and

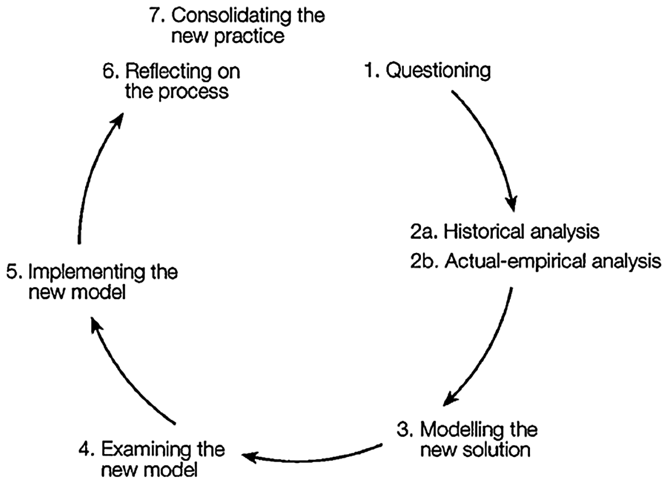


Fig. 6. Expansive cycle of learning actions.

implementing and *generalizing* the final concept of the activity – in CMC terms, internalizing the common ground committed through embodying the common ground in their actions [7–10]. Finally, as seen in the first learning act, the residual of unresolved internalizing knowledge will initiate another learning cycle and the process repeats itself again indefinitely.

Boundary objects then, in their myriad forms, are bridges between different communities of practice [11] or even social worlds [12]. Via their plasticity of meaning they translate ideas, viewpoints, and values across otherwise difficult to traverse organizational boundaries. Social media for knowledge creation as described in CMC can be viewed as boundary objects. In [11], Wenger elaborates on the idea of boundary crossing in the context of communities of practice. He argues that an established practice not only creates separation from others, i.e., boundaries, but also develops opportunities to interact with people in other practices. He explains that boundary objects help bridge different communities of practice, and that they can be digital documents or spaces, standardized methods or concepts. In [13], we have discussed how the social media has been integrated into a blended learning environment that indeed produce Expansive learning cycle as anticipated by Wenger.

In a more recent development of SECI research [6], the distinction between Boom-up and Slip-down SECI processes is made. By foregrounding the ontological dimension for SECI, it is observed the barriers to be overcome are similar to the phenomenon in a boundary crossing activity along the hierarchy of an organization. As argued above, it is clear that Expansive Learning is a self-sustaining activity system where the (intermediate) end of a series of learning activities would produce, by itself, the impetus that engenders a new series of learning activities, and result in a self-sustaining learning cycle. However, such self-sustaining nature of knowledge processes remain un-explained. This is because the end of Boom-up, as explained in [6], is commonly accepted knowledge, typically forming the basis of an innovative product that exploits its commercial benefits. At the same time, Slip-down is explained as a detour that is to be overcome for assuming the Boom-up course of activities. The uni-directional process from local to global truth, along the ontological dimension, does not necessarily need to repeat itself in cycles. For knowledge creation process to be self-sustaining, what more in the processes needs to be brought to bear? With a careful examination, it is realized that a similar process to Slip-down needs can be identified. With reference to Fig. 4, Boom-up is clock-wise and global/collective-wise, activity while Slip-down is counter-clock-wise and local/individual wise, activity. However, the 3-dimensional CMC, as it is realized, allows us to hypothesize two activities that are symmetric to Boom-up and Slip-down, called here *Boil-up* and *Percolate-down*, respectively. Boil-up, a counter-clock-wise and global/collective-wise activity, representing the situation where a failed Internalization where revokes from the ground against the top arise. On the other hand, Percolate-down, a clock-wise and local/individual-wise activating, represents a successful Internalization where the ground indeed engages in developing the personalized practice of the accepted norm, which may produce genuine issues that would engender another cycle of learning activity. This can be summarize in Table 2.

Table 2. Symestrical knowledge creation process types

	Global/Collective-wise	Local/Individual-wise
Clock-wise	Boom-up	Percolate-down
Counter-clock-wise	Boil-up	Slip-down

The above observation points to the pervading tension between tacit-explicit, local-global, and consenses-discensus knowledge conversion processes. This tension provides a theoretical explanation for the self-sustaining nature underling all knowledge creation process. Thus, the social media scaffold in CMC is to function as boundary objects created to bridge the stages of knowledge creation processes as studied in [13].

4 Concluding Remarks and Future Research

This paper discusses Cascading Modes of Communication (CMC) for knowledge creation in social media based on two concepts: Multi-Stakeholder Learning Dialogue (MSLD) and Expansive Learning from the Epistemological, the Ontological and the Communicative perspectives. By establishing mapping between MSLD and SECI, and then Expansive Learning and SECI, it is shown that the three 3-dimensional CMC can serve to unify the learning activities discussed in these theories.

With the 3-dimesional CMC, it is further demonstrated: the two SECI types Boom-up and Slip-down in [6] can be paralleled by two symmetric SECI processes called Boil-up and Percolate-down, respectively. Most importantly, with Percolate-down, it provides a potential opening that SECI cycle can be continued; thus explain the self-sustaining nature of all knowledge creation processes.

This paper is the first attempt to explore how 3-dimensional CMC as a unifying theory can surface potential issues in previous theories; at the same time, provide added theoretical explanatory power for re-analyzing knowledge creation process with new proposal. Nonetheless, it remains to be demonstrated the preliminary results discussed in this paper can be applied in a comprehensive analysis of the cases previously studies. Such application would be able to expose potential limitations of CMC as a theoretical framework.

Another important aspect of CMC is that it formulates a methodology to deploy social media as boundary objects for scaffolding boundary-crossing activities. It remains to be seen, after a preliminary case study [13], whether CMC can prove to be a practical tool to sustain robust expansive learning in activity systems.

Acknowledgements. The author gratefully acknowledges the contribution of Professor Lorna Uden for her discussion and comments, especially in the area of Activity Theory and Expansive Learning cycles.

References

1. Wu, P.H., Uden, L.: Knowledge creation process as communication - connecting SECI and activity theory via cascading modes of communication. In: Proceedings of KMO 2014, pp. 403–412 (2014)
2. Eriksson, O.: A generic communication model based on Habermas and Searle's versions of speech act theory. Listener **1**, 5 (1999)
3. Searle, J.R.: Speech Acts: An Essay in the Philosophy of Language, vol. 626. Cambridge University Press, Cambridge (1969)
4. Deetz, S.: Crossroads-describing differences in approaches to organization science: rethinking Burrell and Morgan and their legacy. Organ. Sci. **7**(2), 191–207 (1996)
5. Calton, J.M., Payne, S.L.: Coping With paradox multistakeholder learning dialogue as a pluralist sensemaking process for addressing messy problems. Bus. Soc. **42**(1), 7–42 (2003)
6. Wu, Y., Senoo, D., Magnier-Watanabe, R.: Diagnosis for organizational knowledge creation: an ontological shift SECI model. J. Knowl. Manag. **14**(6), 791–810 (2010)
7. Engeström, Y.: Learning by Expanding: An Activity-Theoretical Approach to Developmental Research. Orienta-Konsultit, Helsinki (1987)
8. Engstrom, Y.: Innovative learning in work teams: analyzing cycles of knowledge creation in practice. In: Perspectives on Activity Theory, p. 377. Cambridge University Press, Cambridge (1999a)
9. Engeström, Y.: Expansive visibilization of work: an activity-theoretical perspective. Comput. Support. Coop. Work (CSCW) **8**(1–2), 63–93 (1999)
10. Engestrom, Y.: Activity theory as a framework for analyzing and redesigning work. Ergonomics **43**(7), 960–974 (2000)
11. Wenger, E.: Communities of Practice: Learning, Meaning, and Identity. Cambridge University Press, Cambridge (1998)
12. Becky, B.A.: Sharing meaning across occupational communities: the transformation of knowledge on a production floor. Organ. Sci. **14**, 312–330 (2003)
13. Horng-Jyh, P.W.: Knowledge and community formation via cascading modes of communication with a case study and research design. In: Uden, L., Wang, L.S.L., Rodríguez, J.M.C., Yang, H.-C., Ting, I.-H, (eds.) The 8th International Conference of Knowledge Management in Organizations, pp. 189–204. Springer, Netherlands (2004)

The Unspoken Requirements - Eliciting Tacit Knowledge as Building Blocks for Knowledge Management Systems

Sanath Sukumaran[1(✉)] and Kanchana Chandran[2]

[1] Taylor's University, School of Computing and IT, 1, Jalan Taylor's, Taylor's University, 47500 Subang Jaya, Malaysia
sanath.sukumaran@taylors.edu.my

[2] Sunway University, No. 5, Jalan Universiti, Bandar Sunway, 46150 Petaling Jaya, Selangor, Malaysia
kanchanac@sunway.edu.my

Abstract. This conceptual paper sheds light on how analysts deal with the issue of unspoken or unconscious requirements that users might have with regards to Knowledge Management Systems (KMS). Current approaches to requirements elicitation of KMS tend to concentrate mainly on technical aspects, but they ignore human and social influences. How do analysts recognize requirements that users don't or can't tell them about? What approaches have been used and are there (if any) methods that have specifically been developed to identify such requirements? Are KM researchers and practitioners in the area of requirements elicitation aware of the ongoing research in the social and behavioural fields? The paper sheds light on how to improve the way analysts currently obtain requirements from stakeholders. These proposed Knowledge Management – Requirement Elicitation (KMRE) framework is a resource to elicit requirements regarding the human context of a system using a set of analytical techniques and knowledge from Activity Theory and Co-creation. The framework enables collaborative work between requirements engineers, who gather these inputs in the form of software requirements, social practitioners, who provide the knowledge and processes underlying these tools, and the stakeholders, who know the domain and intended application of the projects.

Keywords: Knowledge management (KM) · Knowledge management system (KMS) · Activity theory (AT) · Requirements elicitation · Co-creation

1 Introduction and Background

"Knowledge Management (KM) is the way you manage your organisation, when you understand the value of your knowledge" [1]. In other words, it is the management framework (of roles and accountabilities, processes, technologies and governance) that an organisation puts in place to maximise the value and application of knowledge, and which provide a managed approach to building, developing and retaining know-how, in service of business goals [1].

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 26–40, 2015.
DOI: 10.1007/978-3-319-21009-4_3

Knowledge Management Systems (KMS) refer to any kind of IT system that stores and retrieves knowledge, improves collaboration, locates knowledge sources, mines repositories for hidden knowledge, captures and uses knowledge, or in some other way enhances the KM process [2]. It is a systematic attempt to explicate tacit knowledge (by making tacit knowledge explicit). KMS focuses on discovering knowledge that responds to the changing environment and takes into account tacit knowledge that plays an essential role in an organisation's competitive advantage [3].The notion of KMS as a mere evolution of Information Management is unfounded. KMS today is seen as a highly contextualised source of tacit knowledge driven by socio-economic needs and governed by KM principles.

Typically after a software development project has been commissioned and the planning has begun, there are two important steps that must be taken by a system's analyst before software design can begin. The analyst(s) must elicit requirements from users, customers and other stakeholders and then structure those requirements in a way that can be used for the design specifications and for further follow-up and analysis in future iterations. The practice is referred to as requirements elicitation. Figure 1 below depicts how requirements elicitation fits within the larger context of requirements engineering.

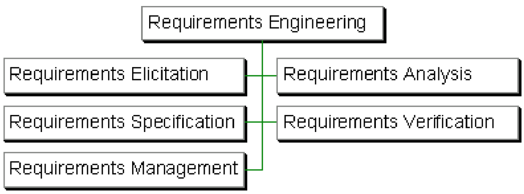


Fig. 1. Requirement engineering activities

The elicitation process involves a set of activities that must allow for communication, prioritisation, negotiation, and collaboration with all relevant stakeholders. Requirements elicitation is a complex exercise that involves activities that are intensely communicative. These activities increase in significance when one considers the "culture gap" or basic semantic differences dividing the problem owning and the problem solving communities when attempting to engage in meaningful dialogue [8]. Requirements elicitation is usually considered as one of the most challenging stages in a software development life cycle since requirements which are incomplete, misunderstood and ambiguous can result in catastrophic software project failures [4, 15].

Michael Polanyi is often recognized as the source of the term "tacit knowledge" [5]. Polanyi observed that "we can know more than we can tell". Polanyi cited a number of examples of "all the things we do effortlessly but cannot explain how we do so" [5]. This stands in contrast to explicit knowledge which can be verbalized and communicated easily (such as the knowledge conveyed in a speech or a book). Tacit knowledge is unconscious and difficult to understand whereas explicit knowledge is conscious and easily explainable. Thus, while it is relatively easy to produce and pass on explicit knowledge, tacit knowledge is conversely difficult to capture.

An organisation with a sound KM strategy in place draws upon itself best practices taking roots from cultural and historical norms, operational excellence and leadership. With a KM strategy in place as pillars of support, a systematic attempt to rollout KMS can be undertaken. A KMS will in turn ensure that KM practices set out under the KM strategy are scalable and sustainable over time.

2 Challenges in Requirements Elicitation in General and KMS in Particular

The requirements elicitation task is difficult with typically multiple stakeholders having different backgrounds, interests and expectations. Figure 2 below illustrate a common scenario;

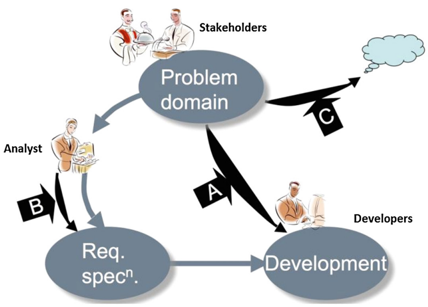


Fig. 2. Problems of requirements elicitation [13]

The arrow labelled "A" shows tacit knowledge that is not made explicit by the analysts but is held by the developers [13]. Arrow "B" shows tacit knowledge of the analysts that are not directly related to the stakeholders but make their way into the requirement specifications anyway [13]. Arrow "C" represents tacit knowledge of the stakeholders that is never explicitly represented to others.

The challenge for an analysts is to figure out how to identify tacit knowledge and make it explicit so that requirements are not misunderstood or misapplied. Being able to identify and document such requirements is difficult and there is no widespread agreement and set of practices for analysts to follow in trying to elicit such knowledge.

KM practices today are multifaceted spanning multi-domains (organisation learning, workflow definition, networking, training and learning, etc.) previously not recognised when it was first coined in the 1990s. If a clear distinction is not made between a KMS and an information systems, a KMS runs the risk of failing and not achieving its intended objectives often centred around business productivity, business intelligence analysis, decision making, efficiency-enhancing and more. It is important to reiterate that for KM systems to be deemed useful, the focus must be towards contextualised tacit knowledge. However, this in itself poses a unique challenge. This is because it is widely recognized that the existence of tacit knowledge poses a unique problem and is a source of difficulty for the knowledge and requirements elicitation process [12–14]. The border between an information system (like POS system, airline, retails system, etc.) and a KMS is diffused and depends on the existence of factors such as an organisation's strategy and goals concerning business and knowledge processes, culture, initiatives, information and communication technologies etc. [2].

The emphasis of tacit knowledge in KMS compared to that of information systems poses yet another challenge and difficulty in building let alone eliciting requirements for such systems. Ironically, the body of knowledge of software requirements engineering as we know today has been tailored for information systems in mind. Knowledge management systems are relatively recent compared to information systems. Although both are fundamentally different, the principle which govern requirements elicitation process for information management type systems are very often than not used to elicit requirements for knowledge management systems. This is indeed a paradox which has challenged KM researchers and developers alike for decades.

It is no surprise that a number of KM implementations today have failed to live up to its expectations and in some cases appears no more than an illustrious off-the-shelf content management system [7]. Requirement elicitation as a body of knowledge is already inundated with various tools and techniques to elicit requirements from various perspectives i.e. human and social elements. While the use of these techniques appears to be somewhat useful for information systems, the same cannot be said for KMS. As Henry Ford said, "If you always do what you've always done, you'll always get what you've always got". Hence, there must be a different approach or perhaps existing approaches are not good enough. Taking the middle path, a better approach would be to augment existing tools and techniques with approaches that will boost the success rates of requirements elicitation for KMS and fulfil the tenets of what KMS has been set out to achieve. Pavónet et al. [9] mentioned yet another challenging perspective, that requirements engineers must pay attention to the human social nature. Significant research must be carried out to perceive how individuals impact the cognitive processes of each other as well as the motivation and emotion of each other. Ineffective communication on the other often result in missing needs, misinterpretation, difference of opinion and reluctance in involvement. A good account of the aforementioned issues were narrated by Werhane et al. [10] who summed up that problems in requirements elicitation involved issues in relation to scope, understanding and volatility - see Fig. 3 below.

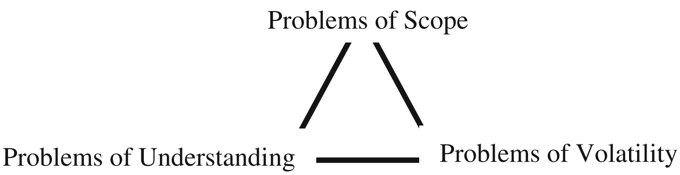


Fig. 3. Problems of requirements elicitation [10]

These unspoken requirements of knowing if enough is enough and the problems of identifying the scope of the system (scope), understanding the needs of the customers/users during the development process (understanding) and the changing requirements over time (volatility) are ongoing challenges faced by analysts since yesteryears. The success of a requirements elicitation exercise relies heavily on understanding and collaboration amongst parties [20]. This is even more so when the requirements elicitation exercise is undertaken for a KM implementation. Since KM relies on codified knowledge residing in the minds of knowledge workers, one must consider various tenets when undertaking KM implementation. Although scholars and researchers alike have pointed out intervening variables like motivation, attention, commitment, creativity, sharing, etc. to better understand and infuse knowledge sharing within KM practice, prevailing research has little light shed as to how requirements for a KMS can be effectively elicited.

This research paper dwells on the challenges to elicit requirements to develop a KMS. The first challenge is the understanding of the characteristics of tacit knowledge (experience and know-how) that goes into a KMS which is an important determinant to ascertain the requirements of the KM system. The second challenge is to determine what constitutes and shapes tacit knowledge. The third challenge in the requirements elicitation process is in finding the necessities of stakeholders for a KMS. To overcome these challenges, requirements engineers must consider using both formal and informal requirement elicitation approaches that are more effective and robust than yesteryears.

3 Co-creation

Co-creation is a management initiative that explores how organisations can improve relationships with their customers (stakeholders) with a view to deliver the best products and services, independent of the idea that it is being done to increase its profits. Co-creation is an active, creative and social process, based on collaboration initiated by the organisation to generate value for its customers [17, 18]. It entails;

- Connections: interactions between people, such as companies and customers, and not just interactions between consumers and products; and
- Collaboration, rather than just involvement

Co-creation requires an ability to engage 'the extended enterprise' by managing across and within customer and supplier of the value creation processes [6]. As stated by Gourlay [13];

> "If stakeholders cannot or will not articulate their objectives thoroughly, the emerging requirements will appear incomplete and possible conflicts will be missed threatening failure or lost opportunities. When critical knowledge, goals, expectations or assumptions of key stakeholders remain hidden or unshared then poor requirements and poor systems are a likely, and costly, consequence."

Operationalising the tenets of co-creation is possible with the DART framework. The DART framework underlines the mechanics on how to operationalise co-creation through four constituent components (building blocks) [11];

– Dialogue represents interactivity between two equal problem solvers, eager to act and to learn.
– Access implies facilitating co-creation by offering the right tools for communication between customers and suppliers.
– Risk assessment is referring to the customers' right to be fully informed about the risks they face from accepting the value proposition.
– Transparency represents symmetrical flow of knowledge between stakeholders.

The DART framework is relevant to this study and the overall discussion of the paper because although the founding authors of DART had in mind customers (from a marketing viewpoint) whom the firm can co-create value, it is no different from the sense of perspective of how internal customer (employees) of the firm behave. It is the same spirit that governs the importance of employees in their company who are prime stakeholders in the knowledge creation process. Hence, for all intents and purposes, in this paper, references will be made to stakeholders of a company as knowledge workers (employees) instead of the commonly used notion of customers.

In understanding co-creation from a KM standpoint, organisations must continuous engage with not only its internal customers (employees) persistently but also with its external customers (where possible) who are both essentially its stakeholders adding to the knowledge creation process. The capacity to engage stakeholders in any KM initiative cannot be underlined enough and has to be planned and carried out in a concerted manner to ensure that knowledge elicited by focusing on creation of value becomes a major source of competitive advantage for an organisation. In effect, if this is exercise is carried out systematically, one would have had KM practices implemented organically across the organisation.

4 Activity Theory

Werhane et al. [10] made an important observation that humans cognitively organize the world and their experiences with the world in their mind (contextualised) and knowledge generated from it may differ from the fact or reality itself. In KM the understanding of the context of a knowledge (environment and conditions that must hold true) is as important as the knowledge itself. This is because if context is lost, then the knowledge cannot be replicated, shared and "consumed" with ease. The term tacit knowledge is used to denote the same idea of contextualised knowledge.

KM practitioners also espouses that knowledge has to be highly contextualised, i.e. parked within a context (an activity). Thus, the benefit of examining knowledge from the viewpoint of an activity is to ensure that a body of knowledge is contextualised. Activity Theory provides a means to achieve the same. In Activity Theory the unit of analysis is an activity directed at an object which motivates activity, giving it a specific direction [21]. Activities are composed of goal-directed actions that must be undertaken to fulfil the object. Actions are conscious. Different actions may be undertaken to meet the same goal and implemented through automatic operations which are also contextualised – see Fig. 4 below. Activity Theory holds that the constituents of activity are not fixed, but can dynamically change as conditions change.

Fig. 4. Basic unit of an activity ([21])

Activity Theory espouses a framework that exemplifies how a knowledge worker (subject) undertakes an activity mediated by explicit knowledge (artifacts) and how the activity fits within a social structure (community) affecting the task being carried out (object) and the resulting outcome. Figure 5 below explains the various tenets of AT that come into play when shaping an outcome for a given activity.

Activity Theory can be used in KM as a tool for understanding and helping with KM design particularly from a requirements elicitation standpoint [22]. A knowledge-intensive activity from a KM standpoint would manifest as a "knowledge activity" whereas components (instruments, subject, rules, community and division of labour) in an Activity System are knowledge sources. The exercise of operationalising Activity Theory in an organisation is relative straight-forward. Firstly, the organisation must undertake an exercise to list out knowledge intensive activities organisation wide. Secondly, one or more knowledge worker is assigned to a knowledge activity. Third, the knowledge worker assigned to a knowledge activity will produce an activity system for that knowledge activity guided by its components listed below in Fig. 5.

Interestingly, since knowledge is highly contextualised, the above mentioned exercise when undertaken by a knowledge worker may yield a different set of insights (knowledge) compared to another knowledge worker for the same knowledge activity. This difference from the viewpoint of Activity Theory is called contradictions which is in effect a rich source of input to the knowledge creation process.

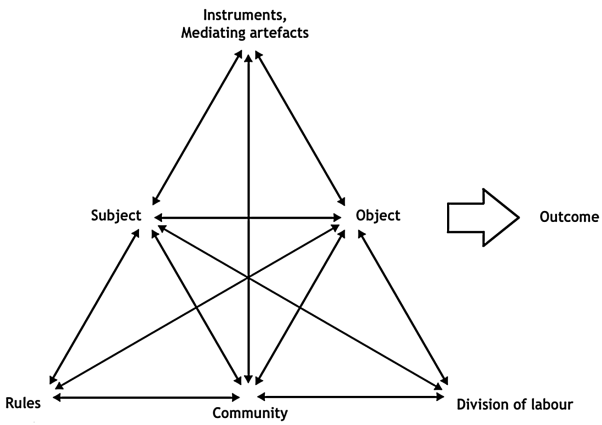


Fig. 5. An Activity System ([21])

Operationalising KM using Activity Theory can be undertaken organisation wide across all its knowledge intensive activities. We have shown how the use of Activity Theory may well be used as a full-fledged knowledge elicitation exercise providing knowledge workers a means to engage in codification of knowledge in a structured and guided manner.

5 KM-RE (Knowledge Management Requirement Elicitation) Framework

An overview of the proposed Knowledge Management Requirements Elicitation Framework (KM-RE) framework is shown in Fig. 6 below;

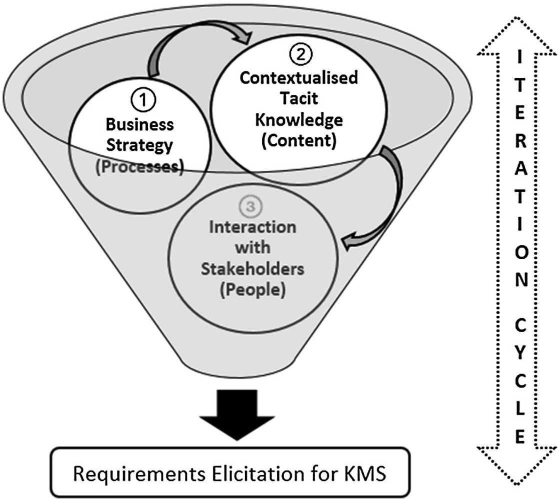


Fig. 6. Knowledge management requirement elicitation (KM-RE) framework

The above framework is outlined through a series of THREE (3) broad steps (as depicted in Fig. 6 above) and outlined further below through a series of inputs and outputs (see Table 1 and the explanation provided);

<u>Step 1: Business Strategy (Processes)</u>
Knowledge Management initiative has always been associated to business strategy. As Milton [16] puts it;

> "Knowledge management is not something you do for its own sake; it's not something to do because "it's a good thing to do". Knowledge Management is something you do in service of your organisational goals. That means that KM needs to be very closely allied with the business strategy."

Table 1. Sources for requirements elicitation – tools/techniques and output

Requirement sources	Tools/techniques	Output
Step 1: Business strategy (processes)	• Which are key strategic business issues/goals that can be significantly impacted by giving people access to knowledge?	• Knowledge assets
	• How can we apply knowledge management to deliver that impact?	• List of knowledge intensive activities
Step 2: Contextualised tacit knowledge (content)	• Knowledge assets-activity mapping	An activity system consists of the following;
		– Instruments
		– Object
		– Subject
		– Rules
		– Community
		– Division of labour
		– Outcome
		• Establishment of knowledge flows between activity systems
Step 3: Interaction with stakeholders (people)	• DART	• For each knowledge activity, the stakeholders will document
		– (D) needs, desires, fears
		– (A)Technology Req.
		– (R) Ownership – benefit (rights, recognition, confidentiality and reward)
		– (T) Knowledge flow

It is important that Step 1 is undertaken first to ensure that priority is established. The following trigger questions can aid in helping prioritise knowledge intensive activities;

- Objective: Identifying knowledge activities (to contextualise knowledge) within an organisation
  Which are the key strategic business issues/goals that can be significantly impacted by giving people access to knowledge?
  Example: Delivering a business proposal, preparing a tender document.
- Objective: To prioritise knowledge activities
  What do I need to know to deliver that impact?
  Example: Getting right access to information and people. Being able to find out the key criteria in preparing, for example, a business proposal or a tender document.

Example of Output:
Knowledge Assets (Know-how) – Business proposal, Tender document.
Knowledge Intensive Activity – Preparing a Business Proposal, Preparing a Tender Document.

Step 2: Contextualised Tacit Knowledge (Content)
Knowledge is not only harder to manage compared to information, it is also highly contextualised. The perennial issue of codification of knowledge is how to explicate tacit knowledge. Activity Theory can be used in this regard to help explicate tacit knowledge. After a knowledge activity has been identified from Step 1, Activity System is created by identifying its components: Instruments, Object, Subject, Rules, Community, Division of Labour and Outcome (see Fig. 5).

Example of Activity System: Activity System for Preparing a Business Proposal Object – Preparing a Business Proposal. It is important to keep tabs of the various elements required under the Business Proposal and coordinating them with various subjects to ensure that it is not only current but also correct.

Instrument – Records of past project proposals, templates, lessons learnt when preparing proposal in the past.

Subject – Client, Project Manager, Project Sponsor, Project Team.

Rules – See Proposal Guide document in the shared folder (Company Records →Proposals). Also look into the company's internal policies when preparing a proposal. See Proposal Preparation Policy in the shared folder (Policy → Policy Guidelines Document).

Community – Stakeholders affected by the proposal. The interest of the client must be considered first. Speaking openly about financial goals. Meeting the CEO to discuss strategic needs to be considered including profit margins, etc.

Division of Labour – Project team involved in the preparation of the proposal. Each member of the team to be delegated a specific task like research and analysis of past figures including ROI, insights into other proposal submission in the past, dealing with client (what protocol to follow), best practices to adhere to, etc.

Outcome – Completed Business Proposal.

It must be noted that there are "cultural and contextual elements" resulting in potential contradictions. For instance, as in the case of preparing a Business Proposal, a knowledge worker may choose to document his or her activity system in a particular way (drawing from his or her experience). However, the approach may not be similar or in agreement with another person (knowledge worker) who may be used to preparing business proposals differently. This draws upon the concept of contradictions (mentioned earlier in Sect. 4 of this paper) which presents an opportunity to augment an activity system over time, i.e. updating of the knowledge base. These contradictions can be due to external (example: client request) or internal reasons (example: change of policy). Whichever the case, tacit knowledge revisions can now take place as and when they occur resulting in a rich ecosystem of knowledge flow within an activity system enriching an organisation's knowledge assets over time.

Step 3: Interaction with Stakeholders (People)
Knowledge management initiatives has to be approached top down. Once high level perspectives in the form of processes and content are mapped out, it is important to

ensure that appropriate buy-ins are in place. Stakeholders of a KMS primarily include an organisation's internal customers, which are very often their employees. What would they need to know do their tasks well? What considerations analysts must take into account to ensure their needs, desires and interests are looked into? The DART (Dialog, Access, Risk-Benefit and Transparency) framework by Li et al. [19] was adapted as follows to address the aforementioned questions;

(D)Dialog	Contribution/Speaking Up (Needs, Desires, Fears) – It includes issues of interest of both the stakeholders and the organisation. It must have clearly defined rules of engagement (defined by means of the organisation's KM policy).
(A)Access	Technology Requirement - Must decide on how best to facilitate the creation and access of knowledge assets with the use of appropriate technology.
(R)Risk-Benefit	A knowledge worker (stakeholder) is responsible towards one or more knowledge activities and has organically maintain his or her knowledge assets over time. The stakeholder runs the 'Risk' of making sure that knowledge is not obsolete. In return the stakeholder shall be bestowed Rights, Recognition, Confidentiality and Reward (subject to the firm's KM Policy) as 'Benefit' [7]. In the spirit of Co-creation, the firm and the stakeholder agree on the definition of these terms.
(T)Transparency	Knowledge Flow – An activity system for a knowledge activity may become an input to one or more activity systems and vice-versa. Figure 7 below shows a hypothetical view of knowledge flows between activity systems depicting precedents and antecedents of an activity system a software system. Such a flow is invariably tied in closely with an organisation's process workflow.

The authors have shown in Step 3 how the DART framework has been adapted to facilitate Co-creation (engagement) amongst stakeholders which is a key element for a successful KM system implementation in particular and for KM initiatives in general.

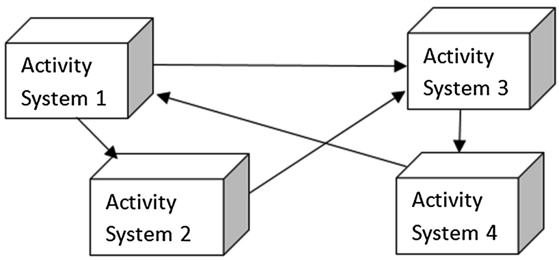


Fig. 7. Knowledge flow between activity systems

The Iteration Cycle

It is important to note that an activity system (see Step 2 under Table 1) is created for all knowledge activities in the business. Hence, the iteration cycle of steps 1 to 3 are repeated for all knowledge activities in the firm starting from knowledge which is of highest priority and critical to the business. For instance, during the initial iteration, the activity system for the knowledge activity of 'Preparing a Business Proposal' is attempted first. This is subsequently followed by the activity system of 'Preparation of a Tender Document'. This then is followed by other knowledge activities in the order of criticality and impact it has on the business. Once this exercise (activity systems) is completed for all the knowledge activities in the firm, an organisation would have had answers of what one needs to know to carry out each and every knowledge activity.

Finally, the proposed KM-RE framework cannot exist in isolation for a holistic requirements elicitation exercise to take place. Hence, the authors espouse that the proposed KM-RE framework ought to be augmented with existing body of knowledge termed as "Requirements Engineering Body of Knowledge". It is also important to note that a KM system can only be commissioned if KM practices are in place from ground up, collectively termed as "KM Environment". Figure 8 below has them all in perspective.

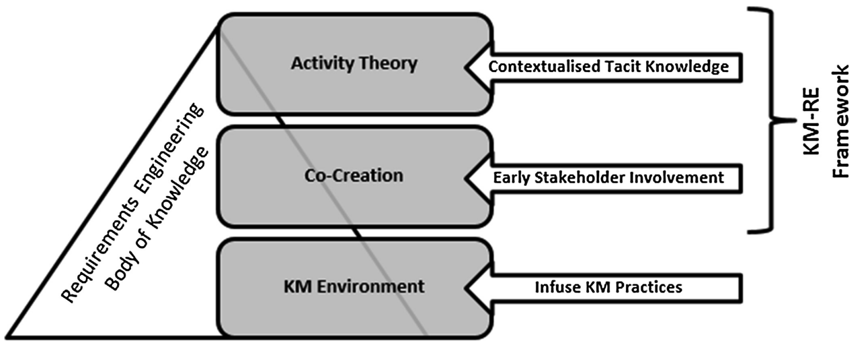


Fig. 8. Requirement sources to undertake RE for KMS (Adapted from [7])

There is another component required for successful KM implementation which is governance which is beyond the scope of this paper. However, since this paper focuses on the aspects of requirements elicitation required to build a KM system, the aspect of governance would not be part of the discussion. Suffice to say that when a KMS is being commissioned, it would have obtained necessary approvals, thus achieving the need for appropriate governance.

Figure 9 below depicts an end-game scenario where activity systems are documented for all knowledge activities in an organisation using a KM system. The resulting effect is a repository of activity systems for an organisation, i.e. knowledge assets. Knowledge assets are also called intellectual capital since it is an amalgam of human, structural, and recorded resourced available to the organisation. Assets reside within the minds of stakeholders and also include physical structures and recorded media. These assets are effectively captured using the THREE (3) Step process outlined in this section.

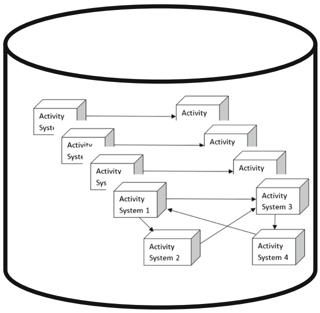


Fig. 9. Knowledge assets based on activity systems

6 Conclusion

Informed, networked, empowered, and active stakeholders are increasingly co-creating value for an organisation, whether consciously or unconsciously. An organisation that capitalises on such a synergy will undoubtedly propel itself forward in its quest towards harvesting its intellectual capital. The human context within which a software system will operate is fundamental for software systems in general and KM systems in particular. Since KM systems are essentially set out to capture human experiences, i.e. tacit knowledge, it is imperative that human, culture and socials paradigms are central to KM initiatives. Such an initiative typically begins with a requirement elicitation exercise in a software development lifecycle. It would, however, be far-fetched to expect an analyst trained in conventional approaches of requirements elicitation to undertake the same for a KMS when the tenets of the former is vastly different from the latter. Even if they do, trained requirement engineers usually have a background in Software Engineering and are not experienced to elicit this kind of information stemming from human and social paradigms expected of a KM system.

The proposed KM-RE framework has its root from well-established theory from social sciences, i.e. Activity Theory and augmented by the management initiative of Co-creation. The KM-RE framework outlined in this position paper is geared towards augmenting existing practices in requirement engineering body of knowledge resulting in requirements specifications specifically for KM systems that are more holistic than yesteryears.

The KM-RE framework also doubles up as a building block in documenting unspoken requirements central to a requirements elicitation exercise taking into account human and social contexts required for a successful KM system implementation. This unspoken requirement which is a symmetrical interaction between an organisation and its stakeholders is undoubtedly one of the most untapped source of rich contextualised tacit knowledge and an important input for requirements elicitation and the ensuing requirements specification for KM systems in the future.

7 Further Research

The main concern in the application of the KM-RE framework is the challenge in elaborating the properties of the framework as it is highly conceptual. This paper integrates several broad subjects and there is scope for further research work with a broader pathway into literature that would complement and deepen the areas discussed. Firstly, operationalising the framework is an interesting endeavour to see if the theoretical appeal of KM-RE framework is met with an adequate attempt at quantitative evaluation. Secondly given that the author was limited by the number of pages, further development of the steps under the framework would bring more theoretical clarity and understanding. Thirdly, since the paper is highly conceptual given that it was set out as a position paper, the challenge of translating the proposed framework into a set of structured guidelines for KM practitioners and researcher alike is a welcome step forward.

References

1. Knoco Ltd. Knowledge Management Reference. Knowledge Management FAQ (2011). http://www.knoco.com/knowledge-management-FAQ.htm. Accessed 27 Feb 2015
2. KMT - An Educational KM Site. Knowledge Management Systems (2013). http://www.knowledge-management-tools.net/knowledge-management-systems.html. Accessed 28 Feb 2015
3. Galandere-zile, I.: Where is the border between an information system and a knowledge management system? Manag. Glob. Transitions **3**(3), 179–196 (2004)
4. Boehm, B.: Software Engineering Economics. Prentice-Hall, Upper Saddle River (1982)
5. Polanyi, M., Sen, A.: The Tacit Dimension, p. 108. Doubleday, New York (1967)
6. Woodruff, R.B., Flint, D.J.: Marketing's service-dominant logic and customer value. In: Lusch, R.F., Vargo, S.L. (eds.) The Service Dominant Logic of Marketing: Dialog, Debate and Directions, pp. 183–195. M.E. Sharpe, Armonk (2006)
7. Sukumaran, S., Chandran, K.: Knowledge management systems - the end of the road?: Position paper. In: Uden, L., Oshee, D.F., Ting, I.-H., Liberona, D. (eds.) KMO 2014. LNBIP, vol. 185, pp. 186–194. Springer, Heidelberg (2014)
8. Van Lamsweerde, A.: Requirements engineering in the year 00: a research perspective. In: Proceedings of the 22nd International Conference on Software Engineering (ICSE-2000), pp. 5–19. ACM Press, New York (2000)
9. Pavón, J., Gomez-Sanz, J.J., Fuentes-Fernández, R.: Part A: systems and humans archive. J. IEEE Trans. Syst. Man Cybern. **39**(2), 282–298 (2009)
10. Werhane, P.H., Hartman, L.P., Moberg, D., Engelehardt, E., Pritchard, M., Parmar, B.: Social constructivism, mental models and problems of obedience. J. Bus. Ethics **100**(1), 103–118 (2011)
11. Prahalad, C.K., Ramaswamy, V.: The Future of Competition: Co-creating Unique Value with Customers. Harvard Business School Press, Boston (2004)
12. Christel, M.G., Kang, K.C.: Issues in requirements elicitation (No. CMU/SEI-92-TR-12). Carnegie-Mellon University Pittsburgh PA Software Engineering Institute (1992). In: Gacitua, R., et al. Making Tacit Requirements Explicit. Second International Workshop on Managing Requirements Knowledge (MaRK 2009) (2010)

13. Gourlay, S.: Towards conceptual clarity for 'tacit knowledge': a review of empirical studies. Knowl. Manag. Res. Pract. **4**(1), 60–69 (2006)
14. Friedrich, W.R., van der Poll, J.A.: Towards a methodology to elicit tacit domain knowledge from users. Interdisc. J. Inf. Knowl. Manag. **2**(1), 179–193 (2007)
15. Knowledge Management FAQ. How does KM tie in with business strategy? (2011) http://www.knoco.com/knowledge-management-FAQ.htm. Accessed 28 Feb 2015
16. Prahalad, C.K., Ramaswamy, V.: Co-creating unique value with customers. Strategy Leadersh. **32**(3), 4–9. ISSN: 1087-8572 (2004) (Emerald Insight)
17. Roser, T., et al.: Co-creation – new pathways to value (an overview). Promise Corporation, LSE Enterprise Report (2009)
18. Prahalad, C.K., Ramaswamy, V.: Co-creation experiences: the next practice in value creation. J. Interac. Mark. **18**(3), 5–14 (2004)
19. Li, K., Dewar, R.G., Pooley, R.J.: Computer-assisted and customer-oriented requirements elicitation. In: Proceedings of the 13th IEEE International Conference on Requirements Engineering (RE 2005), pp. 479–480 (2005)
20. Fernàndez, R.F,, Gómez-Sanz, J.J., Pavón, J.: Understanding the human context in requirements elicitation. Springer-Verlag, Requirements Eng. **15**, 267–283 (2010). doi:10.1007/s00766-009-0087-714, Aug 2009
21. Kaptelinin, V., Nardi, B.A.: Activity theory: basic concepts and applications. In: CHI 1997 Extended Abstracts on Human Factors in Computing Systems, pp. 158–159. ACM, New York, March 1997
22. Uden, L.: Activity theory for designing mobile learning. Int. J. Mob. Learn. Organ. **1**(1), 81–102 (2007)

Modelling Knowledge Management Processes Using Fuzzy Cognitive Maps

Ondrej Prochazka and Petr Hajek(✉)

Faculty of Economics and Administration,
Institute of System Engineering and Informatics, University of Pardubice,
Studentska 84, 532 10 Pardubice, Czech Republic
{st47576, petr.hajek}@upce.cz

Abstract. Both direct and indirect effects have been observed in the relationships between knowledge management processes. Previous attempts have examined only the static nature of these processes. We employ fuzzy cognitive map to show the dynamics in the relationships. The designed model enables concurrent simulation of both direct and indirect effects of knowledge processes on innovation activity. We also show that hyperbolic tangent activation function is the most appropriate one for the modelling of the intrinsic characteristics of knowledge processes. Our results suggest that the indirect effects (via knowledge creation) are stronger in the case of knowledge organization and knowledge acquisition, respectively. The performed sensitivity analysis supports the critical role of the process of knowledge creation.

Keywords: Knowledge management · Knowledge process · Innovation performance · Fuzzy cognitive map · Activation function

1 Introduction

One of the most significant current discussions in knowledge management is the role of knowledge management processes in improving innovation activity in organizations. It is generally accepted that each of knowledge processes (knowledge creation, knowledge sharing, knowledge acquisition and knowledge storage) have a direct impact on organizational innovativeness [1]. In addition, the indirect impacts of the knowledge processes are increasingly attracting attention of both researchers and practitioners. This is because these impacts can be even stronger than the direct ones owing to the effect of knowledge spillovers. However, it is considered difficult to develop methods to measure and isolate the indirect effects.

Recent findings regarding the direct (indirect) effects of knowledge processes on innovation has led to the conclusion that structural equation models can be effectively applied for the analysis. Therefore, several recent studies have produced estimates of the combined effects employing these models for moderated mediation analyses. However, it is only possible to obtain the first-order (linear) relationships between variables. Moreover, for an organization it is also important to see the dynamics in the effects. Fuzzy cognitive maps represent a suitable tool to address these issues. Therefore, this study aims to fill this gap in the literature and develops a model of fuzzy

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 41–50, 2015.
DOI: 10.1007/978-3-319-21009-4_4

cognitive map that: (1) enables concurrent simulation of direct and indirect effects of knowledge processes on innovation activity; (2) simulates the dynamics of the processes using several activation functions; (3) performs sensitivity analysis of the values of variables to show the priorities improving innovation performance in organizations.

In this study, we argue that suitable activation function should be chosen in order to simulate the intrinsic characteristics of knowledge processes. We show that this is one the most critical issue in designing models of fuzzy cognitive maps for this particular domain. Furthermore, we show that the process of knowledge creation is an important vehicle mediating the indirect effect of the remaining knowledge processes on innovation performance.

The remainder of this paper has been divided into four sections. The paper first gives a brief overview of the evidence for the effects of knowledge processes on innovation performance. The next section lays out the theoretical foundations of fuzzy cognitive maps and then it describes the design of a fuzzy cognitive map for the modeling of knowledge processes. Section 4 provides the results of experiments and analyzes the performed simulations. Finally, Sect. 5 concludes this paper and discusses its implications.

2 Knowledge Processes and Innovation Performance

Most of the earlier studies have addressed the impact of selected knowledge processes on innovation performance separately. For example, Deng et al. [2] investigated the effect of absorptive capacity (combining knowledge acquisition, assimilation and application) on innovation performance in IT industry. Xu et al. [3] developed a framework indicating that knowledge processes drive innovation and, in addition, they are interrelated in three different perspectives (physical, human and technological). Wang and Wang [4] report that knowledge sharing (both tacit and explicit) leads to increased financial performance but only through increased innovation speed and quality. Liao and Wu [5] use knowledge management as a single variable including knowledge acquisition, storage and application. They report that knowledge sharing and management commitment mediate the impact of the knowledge management on innovation performance. Darroch [6] also employed structural equation models to show that innovation and organization performances are strongly influenced by the interaction between knowledge acquisition and knowledge sharing. Knowledge creation is reported to be a mediator between knowledge acquisition and innovation by Matusik and Heeley [7]. The mediation role of knowledge creation has been supported by further studies, see [1] or [8] for details.

Taken together, recent studies have considered both direct and indirect impacts of knowledge processes on innovation. Thus, complex models have been proposed to model the mediated relationships. These findings have provided a strong support for further experiments with the dynamics of these relationships. In our further considerations, we follow the methodology proposed by Andreeva and Kianto [1]. In this methodology, four knowledge processes were considered:

(1) Knowledge creation to estimate the frequency of new idea development;
(2) Knowledge storage to measure the intensity of storage and documentation of both tacit and explicit knowledge and the scope of knowledge repositories;
(3) Knowledge sharing to evaluate both vertical and horizontal knowledge sharing within the organization;
(4) Knowledge acquisition to provide information on the frequency of interactions with external environment.

The model developed by Andreeva and Kianto [1] is based on the data analysis of 261 companies from three countries: Finland (38 %), China (33 %) and Russia (29 %). The surveyed companies were with 50 or more employees (mostly between 50 and 500) and represented both production (63 %) and service industries (37 %) with different growth rates. Note that although our design of the fuzzy cognitive map is based on this model of knowledge processes, including the size and direction of the effects, alternative models can be developed accordingly. In fact, there are several approaches to develop fuzzy cognitive maps, either based on expert knowledge or real-life data. These will be introduced in the next section.

3 Fuzzy Cognitive Maps

Robert Axelrod introduced cognitive maps in 1970s [9]. It was used for representing social scientific knowledge. Cognitive maps are signed digraphs consisting of nodes and edges. Nodes represent concepts, which are connected by edges. Each edge represents causal connection between two concepts. Axelrod suggested two types of edges – positive and negative. Positive edge means that if the first node is increasing then the second node is increasing too. Negative edge causes decrease in the second node in the same situation. Cognitive map can be represented by a matrix, where the values of edges are as follows: positive edges (+1) and negative edges (−1).

Cognitive maps describe concepts and relationship between them. Due to fuzziness of knowledge especially in social domain, cognitive maps can be extended by fuzzy logic. By combining cognitive maps and fuzzy logic causality can be described with a better understanding [10]. More specifically, this causality may be represented by a fuzzy relation on causal concepts. Then, a fuzzy cognitive map is a fuzzy signed oriented graph with a feedback. Fuzzy weights with positive/negative signs stand for causal relationships. For the above mentioned features, fuzzy cognitive maps have recently been applied in various domains, including medicine [11], banking [12], mechanics [13] or software engineering [14], see [15] for a review.

The value of node s_i (activation degree – concept value) is calculated as the sum of the current value plus the sum of all incoming edges multiplied with the values of preceding nodes. Then, a threshold function is applied. A directed edge weight w_{ij} between nodes C_i and C_j shows on how much C_i causes C_j. The edge weight w_{ij} lies in the fuzzy causal interval $[-1, 1]$, where $w_{ij} > 0$ denotes a positive causality, $w_{ij} < 0$ refers to a negative causality and $w_{ij} = 0$ indicates no causality. For N concepts C_i, $i = 1, 2, \ldots, N$, an $N \times N$ weighted (adjacency) matrix $\mathbf{W}$ can be constructed. The concept value s_i for each concept C_i is calculated as follows:

$$s_i^{k+1} = f(s_i^k + \sum_{\substack{j=1 \\ j \neq i}}^{N} s_j^k \times w_{ji}), \tag{1}$$

where s_i and s_j stand for concepts C_i and C_j, respectively, k denotes iteration, and f is a threshold function.

Figure 1 shows the design of the fuzzy cognitive map representing causality in knowledge management processes. This design adopts the model of knowledge processes developed by Andreeva and Kianto [1], introducing five concepts: C_1 knowledge organization, C_2 knowledge creation, C_3 knowledge sharing, C_4 knowledge acquisition, and C_5 knowledge application – innovation. Note that all edge weights w_{ij} are greater than zero, this is only positive causalities exist in this fuzzy cognitive map. Both the weights and the directions of edges were also adopted from the conceptual model proposed by Andreeva and Kianto [1].

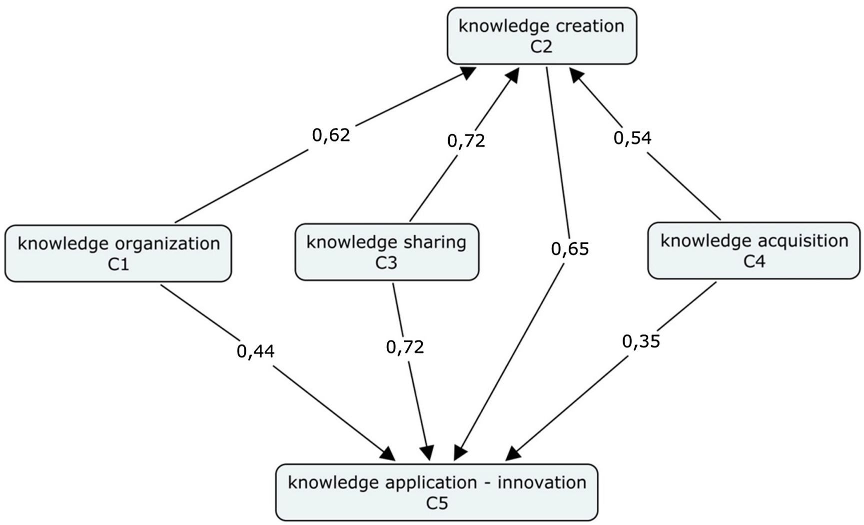


Fig. 1. Fuzzy cognitive map for the modeling of knowledge processes.

This model is based on the assumption that knowledge application depends on both direct and indirect effects. The indirect effects are mediated via the knowledge creation concept. Note that when applying the indirect-effect operator I (minimum) and total-effect operator T (maximum) as defined by Papageorgiou [10], the direct and indirect effects can be compared. For example, knowledge organization affects knowledge application in one direct way (w_{15} = 0.44) and one indirect way. The indirect-effect operators can be then defined as I_1 = min{0.62, 0.65}. Then, the total-effect operator is defined as T = max{0.44, 0.62} = 0.62. In this case, the indirect effect is stronger than the direct effects. The same finding can also be observed for knowledge acquisition. In contrast, the direct effect of knowledge sharing on knowledge application is stronger than the indirect one.

The designed map can also be represented by a weighted (adjacency) matrix **W** (Table 1).

Table 1. Weighted matrix of the fuzzy cognitive map for knowledge processes.

	C_1	C_2	C_3	C_4	C_5
C_1	0.00	0.62	0.00	0.00	0.44
C_2	0.00	0.00	0.00	0.00	0.65
C_3	0.00	0.72	0.00	0.00	0.72
C_4	0.00	0.54	0.00	0.00	0.35
C_5	0.00	0.00	0.00	0.00	0.00

4 Experimental Results

To simulate the evolution of the knowledge management processes defined by concepts C_1, C_2, …, C_5 we used Eq. (1) with $k = 21$ iterations. We set the initial values ($k = 0$) of the concepts arbitrarily to $s_1^0 = 0.2$, $s_2^0 = 0.3$, $s_3^0 = 0.1$, $s_4^0 = 0.5$, and $s_5^0 = 0.7$. In other words, knowledge organization was set to low value, knowledge creation to medium low value, knowledge sharing to very low value, knowledge acquisition to medium value, and knowledge application to medium high value. This setting followed the focus of the research in [1], where developing and emerging economies were predominantly studied. These companies usually rely on knowledge acquisition rather than knowledge creation but, on the other hand, are effective in knowledge application.

In the first set of experiments, we examined the role of activation functions f. We tested linear

$$\begin{aligned} f(s_i) &= 0, s_i < 0 \\ f(s_i) &= s_i, 0 \leq s_i \leq 1 \\ f(s_i) &= 1, s_i > 1 \end{aligned} \tag{2}$$

sigmoid

$$f(s_i) = \frac{1}{1 + e^{-s_i}} \tag{3}$$

and hyperbolic tangent function

$$f(s_i) = \frac{\sin H(s_i)}{\cos H(s_i)} \tag{4}$$

The three activation functions represent the state-of-the-art in FCM modeling [16]. Figure 2 shows that using sigmoid activation function leads to the continuous growth of all knowledge processes. The stable concept values (s_i^7 = 0.659, i = {1,3,4}) of knowledge processes C_1, C_3 and C_4 were achieved after k = 7 iterations.

Although sigmoid activation function allows reaching scenarios far from the extreme values (see [16] for details), note that the continuous growth in knowledge processes does not correspond to reality. Knowledge processes need incentives to grow. Usually knowledge management initiatives have to be implemented to support their growth. A similar issue is associated with the linear activation function where the stability is achieved only after k = 3 iterations. Another disadvantage of sigmoid activation functions is the need for an extensive number of interactions to reach the stable scenario. It has been demonstrated that it requires the making of the quadruple of hyperbolic tangent function interactions, and the triple of step and threshold linear function interactions [16].

In contrast to sigmoid function, employing hyperbolic tangent function promotes initial increase in C_2 and C_5 owing to the incoming edges. However, as knowledge processes C_1, C_3 and C_4 decrease in time, knowledge creation C_2 and knowledge application C_5 start decreasing after k = 4 and k = 5 iterations, respectively. The decrease is stronger for higher initial concept values. Note that hyperbolic tangent function gives the values of concepts in the range [−1, 1], while sigmoid function operates on the fuzzy interval [0, 1].

Taken together, the choice of activation function plays an important role in the behavior of fuzzy cognitive maps. Without applying activation functions in Eq. (1), the behavior of the fuzzy cognitive map would be heavily dependent on the initial values of concepts [17]. In terms of convergence, all activations functions used in the fuzzy cognitive map reasoning led to convergence to a fixed equilibrium point although, generally, the simulation of fuzzy cognitive maps can also lead to a limited cycle or chaotic behavior respectively [17].

Following the above considerations, we used hyperbolic tangent activation function in subsequent sensitivity (what-if) analysis (Fig. 3). The what-if analysis plays an important role in the decision-making process because future implications of various scenarios can be simulated.

To examine the effect of change in edge weights we decreased the weights to 50 %, i.e. w_{12} = 0.31, w_{15} = 0.22, etc. We selected this value to simulate a significant decrease in knowledge management flow. This effect may be caused by worsening key knowledge management factors such as senior management support, knowledge-friendly culture or a clear strategy for managing knowledge [18]. Figure 3a, b show the original setting vs. edge weights decrease. It is obvious that the evolution of concepts C_1, C_3 and C_4, respectively their concept values, remained the same. In contrast, the maximum values of C_2 and C_5 were achieved after k = 6 iterations. Although the edge weights were decreased to 50 %, the concept values of C_2 and C_5 decreased to no more than 75.7 % and 85.4 %, respectively. This finding suggests that the direction of the edges is more important than their actual values.

Figure 3c, d show the sensitivity of the fuzzy cognitive map to the change in concept values. We tested both increase and decrease of the concept values. Figure 3c shows that increasing the concept values leads to only limited change in the evolution patterns of C_2 and C_5. The decrease of the concept values by 0.1, on the other hand, leads to substantial changes in concepts C_2 and C_5. More specifically, their increase is much stronger in this case, although it takes more iterations to achieve maximum

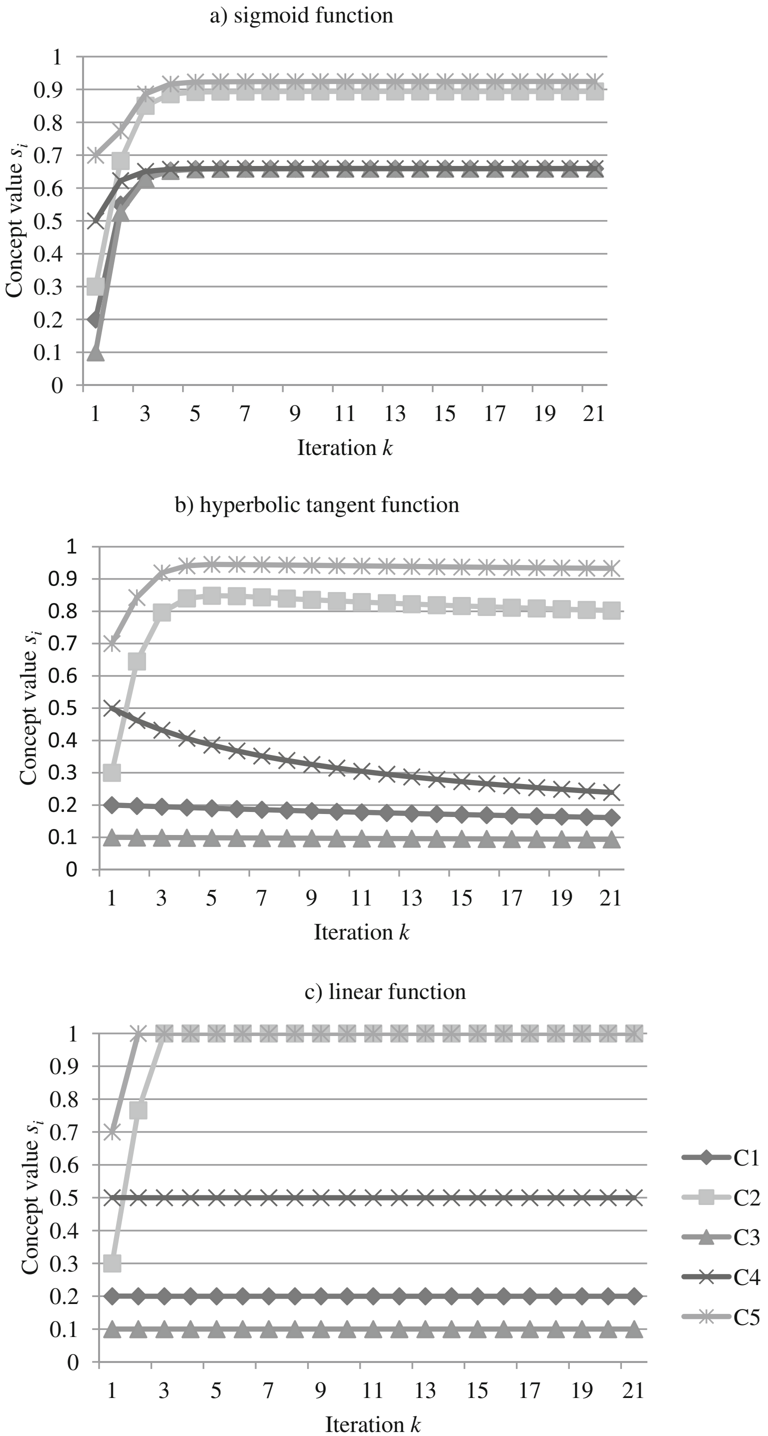


Fig. 2. The effect of various activation functions on concept values, all experiments performed with the same initial concept values and edge weights, (a) sigmoid activation function, (b) hyperbolic tangent activation function, (c) linear activation function.

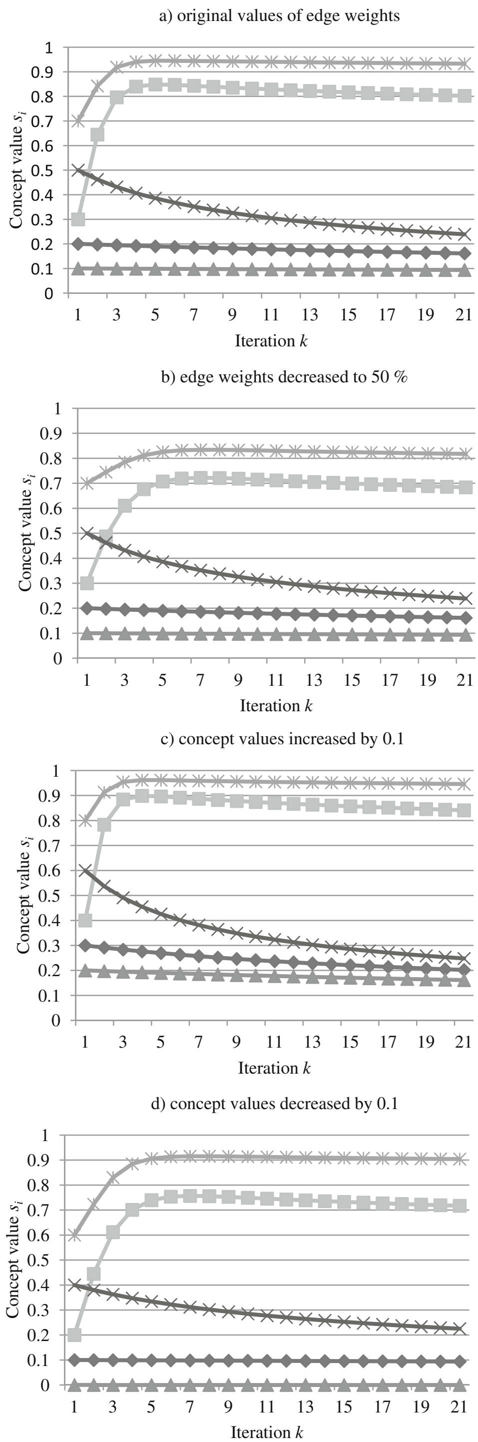


Fig. 3. Hyperbolic tangent function was used to demonstrate the change of input parameters. Graphs (a) and (b) show the change of the edge weights, graphs (c) and (d) show the changes of concept values by 0.1. The concepts C_1 to C_5 correspond to those displayed in Fig. 2.

concept values. The same edge weights were set as in the original setting of the fuzzy cognitive map.

5 Conclusion

In this paper we have outlined the conceptual model of knowledge management processes. This model explains the relationships between concepts represented by knowledge organization, creation, sharing, acquisition and application. The evidence from this study intimates that the relationships in this conceptual model can be effectively simulated using fuzzy cognitive maps.

The findings of this study indicate that the indirect effects, moderated via knowledge creation, may represent a stronger determinant of knowledge application process. The concurrent simulation of the dynamics of knowledge management processes suggests that hyperbolic tangent activation function realistically model the intrinsic character of the studied knowledge processes. The results support the idea that the very existence of indirect effects is the most critical moment in the conceptual model. This effect, no matter how strong the edge weights are, leads to substantial increase in knowledge creation and knowledge application, respectively. The upshot of this is that it is much more important for organizations to support internal knowledge creation through knowledge organization, sharing and acquisition than the direct application of knowledge acquired externally. Therefore, technologies such as brainstorming, concept and mind mapping, decision support applications or training employees on how to develop new ideas or creativity may provide effective knowledge creation tools.

Finally, a number of caveats need to be noted regarding the present study. In our study, we employed the conceptual model which is probably correct for the firms examined in [1] but may be different for various industries and cultural settings. Furthermore, in our experiments we tested only limited amounts of concept values. Further experiments on real-world data are necessary to provide support to our findings. Finally, it may be difficult to exactly evaluate both the concept values and edge weights. We are currently in the process of investigating the possibilities of the generalizations of fuzzy cognitive maps, introducing additional modes of uncertainty. We also envision more complex models incorporating additional internal and external determinants of knowledge management processes, for example at the regional level [19, 20]. This is also important for efficient innovation policies [21, 22].

Acknowledgments. This article was created as a part of the solution of the research task No. SGSFES_2015001, financially supported by Student Grant Competition, and No. 14-02836S, financially supported by the Grant Agency of the Czech Republic.

References

1. Andreeva, T., Kianto, A.: Knowledge processes, knowledge-intensity and innovation: a moderated mediation analysis. J. Knowl. Manag. **15**(6), 1016–1034 (2011)

2. Deng, X., Doll, W.J., Cao, M.: Exploring the absorptive capacity to innovation/productivity link for individual engineers engaged in it enabled work. Inf. Manag. **45**(2), 75–87 (2008)
3. Xu, J., Houssin, R., Caillaud, E., Gardoni, M.: Macro process of knowledge management for continuous innovation. J. Knowl. Manag. **14**(4), 573–591 (2010)
4. Wang, Z., Wang, N.: Knowledge sharing, innovation and firm performance. Expert Syst. Appl. **39**(10), 8899–8908 (2012)
5. Liao, S.H., Wu, C.C.: System perspective of knowledge management, organizational learning, and organizational innovation. Expert Syst. Appl. **37**(2), 1096–1103 (2010)
6. Darroch, J.: Knowledge management, innovation and firm performance. J. Knowl. Manag. **9** (3), 101–115 (2005)
7. Matusik, S.F., Heeley, M.B.: Absorptive capacity in the software industry: identifying dimensions that affect knowledge and knowledge creation activities. J. Manag. **31**(4), 549–572 (2005)
8. Smith, K.G., Collins, C.J., Clark, K.D.: Existing knowledge, knowledge creation capability and the rate of new product introduction in high technology firms. Acad. Manag. J. **48**(2), 346–357 (2005)
9. Axelrod, L.: Structure of Decision: The Cognitive Maps of Political Elites. Princeton University Press, Princeton (1976)
10. Kosko, B.: Fuzzy cognitive maps. Int. J. Man Mach. Stud. **24**, 65–75 (1986)
11. Papageorgiou, E.I.: A new methodology for decisions in medical informatics using fuzzy cognitive maps based on fuzzy rule-extraction techniques. Appl. Soft Comput. **11**(1), 500–513 (2011)
12. Xirogiannis, G., Glykas, M., Staikouras, C.: Fuzzy cognitive maps in banking business process performance measurement. In: Glykas, M. (ed.) Fuzzy Cognitive Maps, pp. 16–200. Springer, Berlin (2010)
13. Beena, P., Ganguli, R.: Structural damage detection using fuzzy cognitive maps and Hebbian learning. Appl. Soft Comput. **11**(1), 1014–1020 (2011)
14. Salmeron, J.L., Lopez, C.: Forecasting risk impact on ERP maintenance with augmented fuzzy cognitive maps. IEEE Trans. Softw. Eng. **38**(2), 439–452 (2012)
15. Papageorgiou, E.I.: Review study on fuzzy cognitive maps and their applications during the last decade. In: Glykas, M. (ed.) Business Process Management, pp. 281–298. Springer, Berlin (2013)
16. Bueno, S., Salmeron, J.L.: Benchmarking main activation functions in fuzzy cognitive maps. Expert Syst. Appl. **36**(3), 5221–5229 (2009)
17. Groumpos, P.P., Stylios, C.D.: Modelling supervisory control systems using fuzzy cognitive maps. Chaos, Solitons Fractals **11**(1), 329–336 (2000)
18. Wong, K.Y., Aspinwall, E.: An empirical study of the important factors for knowledge-management adoption in the SME sector. J. Knowl. Manag. **9**(3), 64–82 (2005)
19. Hajek, P., Henriques, R., Hajkova, V.: Visualising components of regional innovation systems using self-organizing maps - evidence from European regions. Technol. Forecast. Soc. Change **84**, 197–214 (2014)
20. Matatkova, K., Stejskal, J.: Descriptive analysis of the regional innovation system-novel method for public administration authorities. Transylvanian Rev. Adm. Sci. **39**, 91–107 (2013)
21. Hajkova, V., Hajek, P.: Efficiency of knowledge bases in urban population and economic growth - evidence from European cities. Cities **40**, 11–22 (2014)
22. Stejskal, J., Hajek, P.: Competitive advantage analysis: a novel method for industrial clusters identification. J. Bus. Econ. Manag. **13**(2), 344–365 (2012)

Performance Measurement of Knowledge Resources Using Fuzzy Logic

Cheng Sheng Lee(✉) and Kuan Yew Wong

Faculty of Mechanical Engineering, Department of Manufacturing and Industrial Engineering, Universiti Teknologi Malaysia (UTM), 81310 Skudai, Malaysia
cyruslee@live.com.my, wongky@mail.fkm.utm.my

Abstract. Evaluating the performance of knowledge resources is not a simple task as it involves many aspects, some of which are bounded by uncertainties and indistinctness. These make it hard to judge or quantify knowledge resources numerically for measurement purposes. Furthermore, data collection for performance measurement seems to be another difficulty faced by many organizations. This paper proposed to measure the performance of knowledge resources using fuzzy logic to cope with the shortcomings of normal measurement methods. The feasibility of using fuzzy logic as an evaluation method was shown with a five-step guideline and an example computed using MATLAB version R2013a.

Keywords: Knowledge resources · Performance measurement · Fuzzy logic

1 Introduction

As competition among companies is becoming more and more intense, it is crucial for them to perform well in every aspect to stay ahead of their competitors. Knowledge has become the prime source of innovativeness and the means to stay competitive [1, 2].

The performance of knowledge resources can be assessed, by evaluating the tacit knowledge and explicit knowledge that a company owns [3–6], depending on how well they meet the standards set out by the company. Researchers have proposed numerous tools and techniques for evaluating knowledge resources such as, Intangible Assets Monitor [3], Skandia Navigator [7], Tobin's q Ratio [8], KP^3 Methodology [9], etc.

Even so, performance evaluation of knowledge resources is not an easy task due to their subjective and intangible nature [10]. Indicators or metrics used for measurement may seem vague and data gathered may contain certain degree of fuzziness where justification will differ from one evaluator to another. Furthermore, without advanced information technological tools or expertise for structured measurement and reporting, it is difficult and time consuming to obtain complete and precise information for the evaluation process.

Therefore, the aim of this paper is to propose an evaluation method using fuzzy logic to cope with the vagueness and uncertainties in the evaluation process, dealing with incomplete and fuzzy data. When exact numerical values cannot be given, it is sensible to use linguistic assessment for measurement [11]. The proposed method is not to replace the current classical measurement methods, instead it is an alternative for

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 51–59, 2015.
DOI: 10.1007/978-3-319-21009-4_5

companies that do not have advanced technological tools to aid their data collection process for evaluation purposes. This paper will briefly discuss about fuzzy logic and its applicability as a measurement method, followed by an example.

2 Fuzzy Logic

The concept of fuzzy logic was created by Zadeh [12]. Its primary application is as a control system, but currently it has been expanded to many fields such as, image processing, artificial intelligence, management, etc.

The advantage of fuzzy logic is that it is conceptually easy to understand as the mathematical concepts behind fuzzy reasoning are very simple. It is tolerant of imprecise data and it does not require precise input information [13]. It provides a simple way to arrive at a definite conclusion based upon vague, ambiguous, imprecise, noisy, or missing input information. Therefore, it is suitable to be used as an evaluation system as the data collected for evaluation are not always precise and complete.

Another advantage of fuzzy logic is that fuzzy if-then rules resemble human thinking as they are built based on human experience [14]. With this, knowledge and experience from experts can be incorporated into the evaluation system, giving it the ability to assess the performance of a company's knowledge resources as how an expert evaluator would.

The fuzzy logic methodology involves five steps [15]. Steps 1–4 are also known as a fuzzy inference system where a given input is transformed into an output using membership functions, fuzzy logic operators, and if-then rules. There are two types of fuzzy inference systems: Mamdani and Sugeno. In this paper, only Mamdani's max–min inference mechanism will be discussed as it is the most commonly used inference mechanism to produce fuzzy sets for defuzzification [15].

2.1 Fuzzification of Inputs

In this step, inputs are converted into membership values or degrees of membership between 0 and 1 through a membership function in a fuzzy set. There are a few membership functions that can be used such as triangular, trapezoidal, Gaussian curve, generalized bell, and so on. Triangular membership function is commonly used because of its simplicity and easy computation compared to the others [15].

2.2 Application of Fuzzy Operators

The input to a fuzzy operator can be two or more membership values from the fuzzified input variables, whereas the output is a single truth value. Once the inputs have been fuzzified, the degree to which each part of the antecedent has been satisfied for each rule will be known. If the antecedent of a given rule has more than one part, a fuzzy operator is applied to resolve the antecedent to a single number between 0 and 1, which is the degree of support for the rule. This number will then be applied to the output function. There are three types of fuzzy operators: AND, OR, and NOT. The fuzzy

AND operator selects the minimum value, OR operator selects the maximum value, and NOT operator is the fuzzy complement.

2.3 Application of the Implication Method

A consequent is a fuzzy set represented by a membership function, which weights appropriately the linguistic characteristics that are attributed to it. The consequent is reshaped using a function associated with the antecedent. The input for the implication process is a single number given by the antecedent, and the output is a fuzzy set. Implication is implemented for each rule. In Mamdani's max–min mechanism, implication is modeled by means of the minimum (AND) operator [15].

2.4 Aggregation of All Outputs

As decisions or final results are based on the testing of all of the rules in a fuzzy inference system, the rules must be combined in order to make a decision. Aggregation is the process of combining all the fuzzy sets that represent the outputs of each rule into a single fuzzy set. In Mamdani's max–min mechanism, the outputs of each rule are combined using the maximum (OR) operator [15].

2.5 Defuzzification

In this step, the aggregated output fuzzy set that encompasses a range of output values is defuzzified into a single crisp value which in this case will be the performance value of knowledge resources. There are a few defuzzification methods, such as centroid, middle of maximum, largest of maximum, and smallest of maximum. The most popular defuzzification method is the centroid calculation, which returns the center of area under the curve [15].

3 Knowledge Resources Performance Measurement Example

This section demonstrates an example on evaluating the performance of knowledge resources using fuzzy logic. A company's knowledge resources can be assessed based on three constructs: human capital (tacit knowledge), knowledge capital (explicit knowledge), and intellectual property [6], as shown in Table 1 with some examples of metrics.

Human capital refers to the employees of an organization as they are the vessel of knowledge. It can be evaluated based on employees' education level and working experience. People usually obtain knowledge through education [16], while working experience is something that accumulates and builds up over time [17].

Knowledge capital refers to the explicit knowledge that a company owns. It can be evaluated based on the quantity and quality of knowledge that is stored in a company's database or knowledge repository.

Table 1. Constructs and metrics for knowledge resources

Constructs	Metrics
Human capital (tacit knowledge)	a_1. Employees' education level
	a_2. Years of experience of employees in the profession
Knowledge capital (explicit knowledge)	b_1. Amount of knowledge stored in database
	b_2. Quality of knowledge stored in database
Intellectual property	c_1. Number of intellectual properties owned
	c_2. Revenue generated from intellectual properties

Intellectual property is another important knowledge asset for an organization [18], as companies are using it to generate income [19, 20]. A company's intellectual property includes its invention, patent, trademark, industrial design, trade secret, and copyright [21].

Each construct has two metrics which are the input variables of the fuzzy logic inference system. Each input variable has five triangular membership functions. As an example, the membership functions of the human capital's metrics a_1 and a_2 are shown in Figs. 1 and 2. The fuzzy sets of the input variables are similar for all six metrics, as shown in Table 2. They are the same for the output (performance result) as shown in Fig. 3 and Table 3. The if-then rules of metrics a_1 and a_2 are shown in Table 4. The rules are the same for b_1, b_2, c_1, and c_2 as well. The surface viewer of the proposed fuzzy system for human capital performance evaluation is shown in Fig. 4.

This section shows a simple evaluation process by using MATLAB version R2013a to apply the proposed fuzzy system for knowledge resources performance measurement. For this purpose, the following linguistic terms and their respective input scores are used: none = 0, extremely low = 10, very low = 20, low = 30, slightly

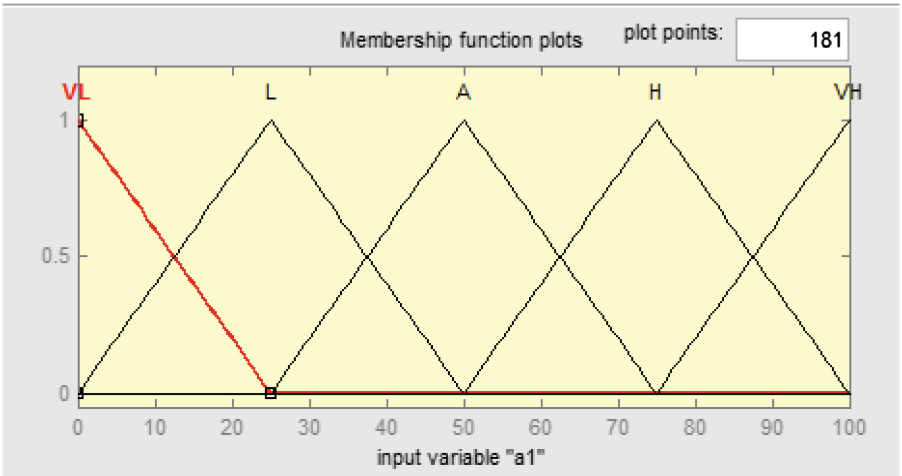


Fig. 1. Membership function of input a_1

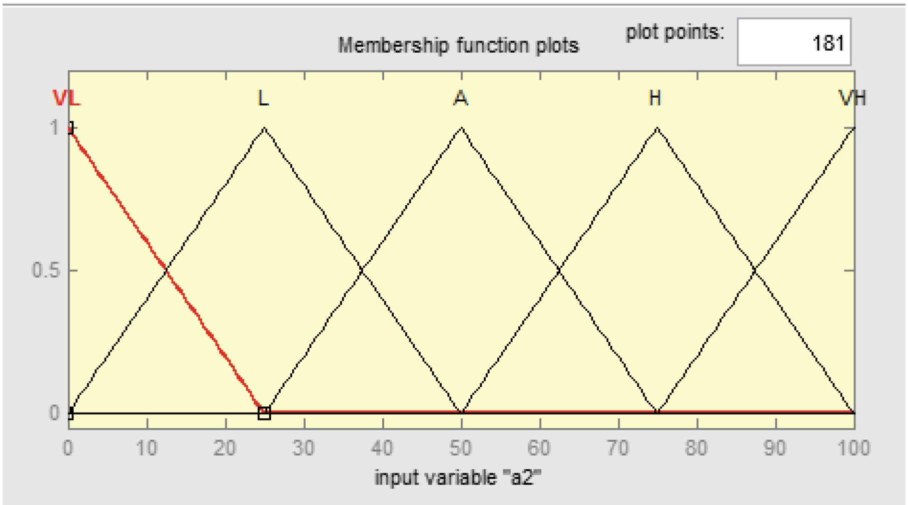


Fig. 2. Membership function of input a_2

Table 2. Fuzzy set of a_1, a_2, b_1, b_2, c_1, and c_2

Linguistic variables	Interval
Very low (VL)	(0,0,25)
Low (L)	(0,25,50)
Average (A)	(25,50,75)
High (H)	(50,75,100)
Very high (VH)	(75,100,100)

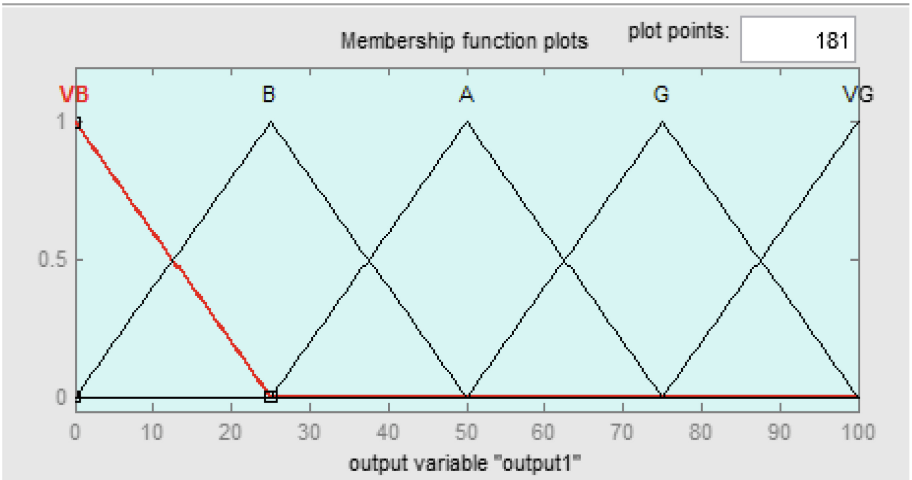


Fig. 3. Membership function of output (performance result)

Table 3. Fuzzy set of output (performance result)

Linguistic variables	Interval
Very bad (VB)	(0,0,25)
Bad (B)	(0,25,50)
Average (A)	(25,50,75)
Good (G)	(50,75,100)
Very good (VG)	(75,100,100)

Table 4. Inference rules of a_1 and a_2

	a_2					
a_1		VL	L	A	H	VH
	VL	VB	VB	B	B	A
	L	VB	B	B	A	G
	A	B	B	A	G	G
	H	B	A	G	G	VG
	VH	A	G	G	VG	VG

low = 40, average = 50, slightly high = 60, high = 70, very high = 80, extremely high = 90, perfect = 100. Assume that the input scores for (a_1, a_2), (b_1, b_2), and (c_1, c_2) are (70, 60), (40, 80), and (20, 60). The computation is based on Mamdani's max–min

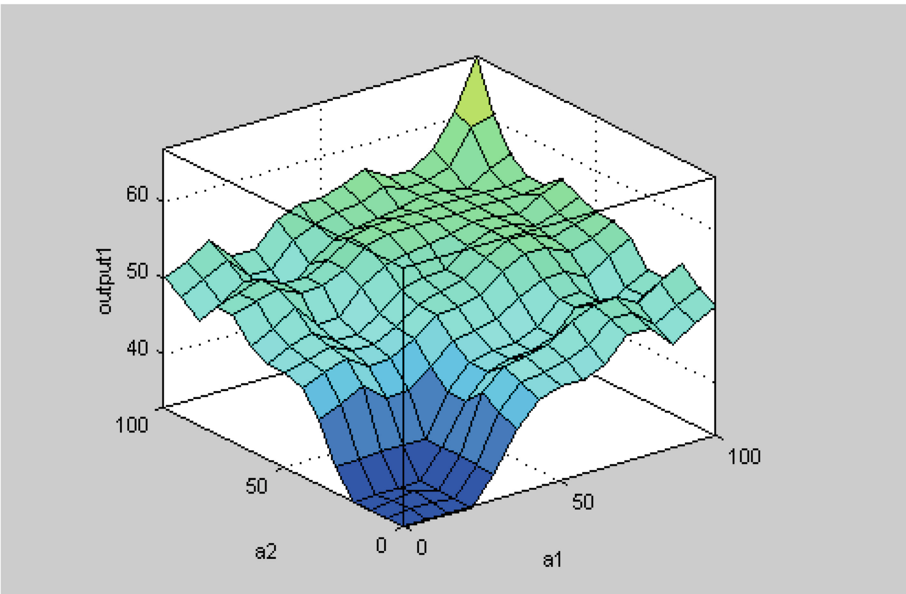


Fig. 4. Surface viewer of KM performance evaluation

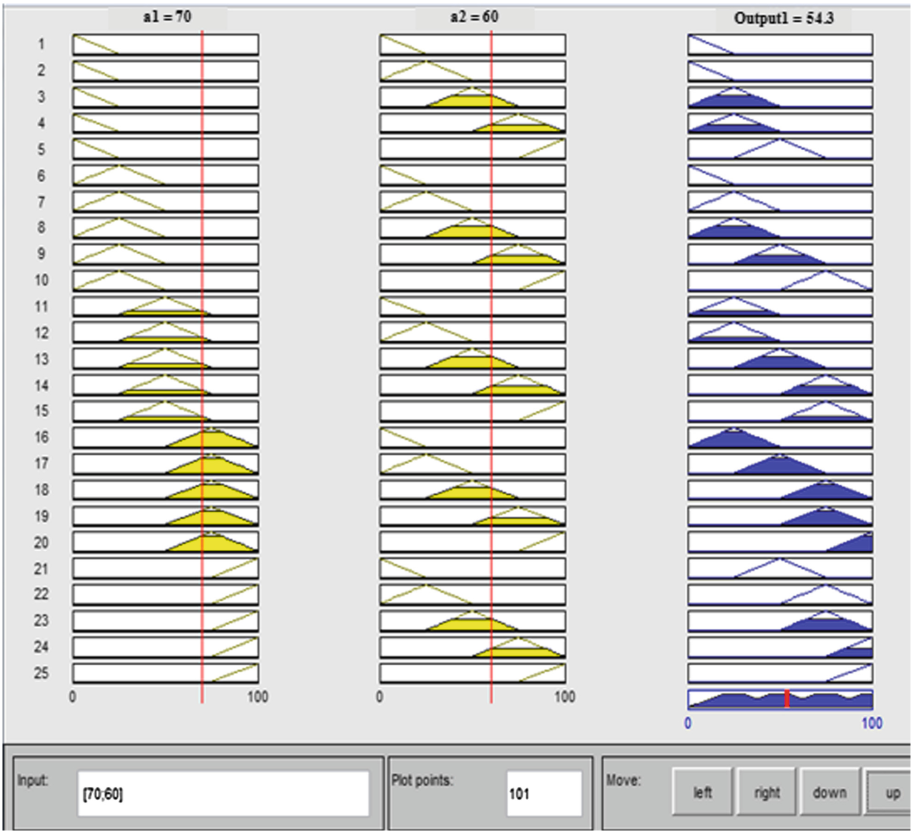


Fig. 5. Active rules and performance value of human capital

inference system. Take the human capital construct as an example. Firstly, each input score is fuzzified by the triangular membership functions. The OR fuzzy operator then selects the maximum value for each rule, followed by the implication step where each rule is modeled by the minimum (AND) operator. The outputs of each rule are then aggregated using the maximum (OR) operator. Finally, the centroid defuzzification method is applied to obtain the performance value for human capital.

Figure 5 shows the active rules and performance value for human capital which is 54.3 out of 100. The same evaluation process is then carried out to evaluate knowledge capital and intellectual property. The results are shown in Table 5.

Table 5. Evaluation results

Constructs	Input score	Output score
Human capital	(70, 60)	54.3
Knowledge capital	(40, 80)	52.6
Intellectual property	(20, 60)	47.4
Total average score		51.4

4 Discussion

From the given example, the output performance scores for human capital, knowledge capital, and intellectual property are 54.3/100, 52.6/100, and 47.4/100 respectively. The overall performance score for knowledge assets is 51.4/100, which is fair to rate their performance as average.

In this paper, a simple example has been shown involving two metrics for each of the three constructs, whereas in actual application it may involve more metrics and constructs. It is wise not to use a large number of metrics for each construct as the number of fuzzy if-then rules will increase exponentially based on the number of inputs.

In practical, the membership functions and rules can be fine tuned to suit a company's requirement and specification as the standard may differ from one company to another. The membership functions and fuzzy if-then rules are normally built based on the experience and judgment of an expert in evaluating the performance of knowledge resources. In advanced cases, a weight system can be added based on the expert's experience on what is more important compared to the others.

5 Conclusions

The feasibility of fuzzy logic as a measurement method for evaluating knowledge resources performance has been shown. It involves five simple steps: fuzzification of inputs, application of fuzzy operators, application of the implication method, aggregation of all outputs, and finally, defuzzification. Overall, fuzzy logic has the potential to be developed into a robust performance measurement tool, helping organizations in evaluating their knowledge resources performance without much effort and in a timely manner.

By having an effective knowledge resources performance measurement system in place, organizations can keep track on their knowledge assets. In addition, constructive information obtained from the evaluation process can help organizations in their decision making and improvement process. Ultimately, companies can ensure the innovation and development of their knowledge resources, thus obtaining the competitive advantage that they seek.

References

1. Kankanhalli, A., Bernard, T.C.Y., Wei, K.K.: Contributing knowledge to electronic knowledge repositories: an empirical investigation. MIS Q. **29**(1), 113–143 (2005)
2. Mills, A.M., Smith, T.A.: Knowledge management and organizational performance: a decomposed view. J. Knowl. Manag. **15**(1), 156–171 (2011)
3. Sveiby, K.E.: The intangible assets monitor. J. Hum. Resour. Costing Acc. **2**(1), 73–97 (1997)

4. Jennex, M.E., Olfman, L.: Assessing knowledge management success/effectiveness models. In: Proceedings of the 37th Hawaii International Conference on System Sciences. Hawaii, pp. 1–10 (2004)
5. Wu, Y.L., Wang, X., Wu, H.S.: Research on the performance measurement of knowledge management based on principal component analysis. In: Proceedings of the International Workshop on Intelligent Systems and Applications. Wuhan, pp. 1–4 (2009)
6. Wong, K.Y., Tan, L.P., Lee, C.S., Wong, W.P.: Knowledge management performance measurement: measures, approaches, trends and future directions. Inf. Dev. (2013). doi:10.1177/0266666913513278
7. Edvinsson, L.: Developing intellectual capital at Skandia. Long Range Plan. **30**(3), 266–373 (1997)
8. Tobin, D.R.: The Knowledge-Enabled Organization: Moving from "Training" to "Learning" to Meet Business Goals. Amacom, New York (1998)
9. Ahn, J.H., Chang, S.G.: Assessing the contribution of knowledge to business performance: the KP^3 Methodology. Decis. Support Syst. **36**(4), 403–416 (2004)
10. Kluge, J., Stein, W., Licht, T.: Knowledge Unplugged: The McKinsey and Company Global Survey on Knowledge Management. Palgrave, Basingstoke (2001)
11. Herrera, F., Herrera-Viedma, E.: Linguistic decision analysis: steps for solving decision problems under linguistic information. Fuzzy Sets Syst. **115**, 67–82 (2000)
12. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**(3), 338–353 (1965)
13. Rahmani, B., Rafezi, H.: Solving fuzzy logic problems with MATLAB (2010). http://www.scribd.com/doc/33412283/Solving-Fuzzy-Logic-Problems-With-MATLAB. Accessed 5 Feb 2015
14. Godil, S., Shamim, M., Enam, S., Qidwai, U.: Fuzzy logic: a "simple" solution for complexities in neurosciences. Surg. Neurol. Int. **2**(24) (2011). doi:10.4103/2152-7806.77177
15. MathWorks: Fuzzy logic toolbox user's guide (2015). http://www.mathworks.com/help/pdf_doc/fuzzy/fuzzy.pdf. Accessed 5 March 2015
16. Uit Beijerse, R.P.: Questions in knowledge management: defining and conceptualising a phenomenon. J. Knowl. Manag. **3**(2), 94–110 (1999)
17. Handzic, M.: Knowledge management in SMEs: practical guidelines. CACCI J. **1**(1), 21–34 (2006)
18. Davenport, T.H., Prusak, L.: Working Knowledge: How Organizations Manage What They Know. Harvard Business School Press, Boston (1998)
19. Clarke, J., Turner, P.: Global competition and the Australian biotechnology industry: developing a model of SMEs knowledge management strategies. Knowl. Process Manag. **11**(1), 38–46 (2004)
20. Autrey, R.L., Sansing, R.: Licensing intellectual property with self-reported outcomes. J. Acc Auditing Finan. **29**(3), 260–277 (2014)
21. WIPO: WIPO Intellectual Property Handbook. World Intellectual Property Organization Publication (2004)

Assessing the Effect of Knowledge Management Initiatives on Stakeholder Objectives Using Fuzzy TOPSIS

Petr Hajek[1(✉)] and Lucie Mansfeldova[2]

[1] Faculty of Economics and Administration,
Institute of System Engineering and Informatics,
University of Pardubice, Studentska 84, 532 10 Pardubice, Czech Republic
petr.hajek@upce.cz
[2] Faculty of Economics and Administration,
Institute of Business Economics and Management,
University of Pardubice, Studentska 84, 532 10 Pardubice, Czech Republic
st37902@student.upce.cz

Abstract. Comprehensive assessment tools to prioritize KM initiatives are critical for KM adoption. This paper aims to develop a hierarchical fuzzy TOPSIS model to rank KM initiatives in terms of their effect on often contradictory stakeholder objectives. The empirical case study of a medical supplier company demonstrates that this group multi-criteria decision making tool can effectively address both the uncertainty in KM assessment and the different effects of KM processes on individual stakeholder objectives. Thus, it may contribute to the successful KM implementation process.

Keywords: Knowledge management · Assessment · Initiatives · Stakeholder · Objectives · Fuzzy TOPSIS

1 Introduction

Today, knowledge is a vital strategic asset, helping to achieve corporate goals. Knowledge management (KM) provides the tools to support knowledge creation, organization, sharing and application. Thus, competitive advantage can be attained through improved innovation ability, coordination of efforts, responsiveness to market change, etc. [1–3]. Therefore, organizations aim to develop and manage organizational culture and structure that promote effective KM initiatives and use KM tools to optimize knowledge flows.

KM initiatives may be beneficial for several reasons. First, they can boost organizational performance, usually measured in terms of profit or market value. Revenue can be increased by using the knowledge of customers, suppliers and competitors. This knowledge plays a critical role in R&D, production and marketing processes. Thus, external knowledge helps firms in acquiring customers and increasing their loyalty. Additionally, firms become more flexible in responding to change in customer preferences. Second, innovation in production process decreases cost and increases product quality. In other words, KM may decrease production loss, material wasting, working

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 60–70, 2015.
DOI: 10.1007/978-3-319-21009-4_6

capital and production time. Finally, it also may improve the efficiency of supplier networks [4].

The selection of appropriate KM initiatives (activities, tools) has become a critical issue in KM implementation. This complex and organization-specific issue is often performed intuitively, although multi-criteria decision making (MCDM) methods can be employed to both rank the initiatives and select the optimum portfolio of the initiatives [5]. However, a major problem with this kind of application is the mix of often contradictory stakeholder objectives. Therefore, the goal of the MCDM problem is becoming more complex and requires novel hierarchical MCDM methods to be developed. This problem is also associated with uncertainty and vagueness in assessing both the contribution of the initiatives to criteria and the weights of the criteria in stakeholder objectives. Fuzzy logic is an approach that handles this issue effectively [6–10].

This study aims to fill the gap and develops a hierarchical fuzzy TOPSIS model to rank the KM initiatives in terms of their effect on stakeholder objectives. We demonstrate the proposed approach on a case of a medical supplier firm.

The rest of the paper is organized as follows. Section 2 briefly reviews the literature on KM initiatives and criteria. The applied methodology based on hierarchical fuzzy TOPSIS method is introduced in Sect. 3. The case study is performed in Sect. 4 and Sect. 5 discusses the conclusions.

2 Literature Review

The research to date has tended to focus on three components vital for KM assessment, namely KM enablers, KM processes and organizational performance [11]. KM enablers represent such organization (technical and social) mechanisms that intentionally and systematically develop knowledge. KM processes usually include knowledge creation, organization, sharing and application. The impact model can be then seen as an input-process-output model [12]. Thus, KM initiatives affect organizational performance through KM processes. Organizational performance is usually assessed by financial indicators owing to their comparability. However, other dimensions of the performance can also be taken into account, for example by using balanced scorecard.

The effect of KM processes on organizational performance was shown to be significantly positive in several studies [13–15]. Andreeva and Kianto [13] report that each of four KM processes (knowledge creation, knowledge sharing, knowledge acquisition and knowledge organization) has a direct impact on organizational innovativeness. Darroch [14] employed structural models to investigate the effects of knowledge acquisition, knowledge sharing and responsiveness to knowledge on profit and sales. Wang and Wang [15] used a similar model to test the effect of knowledge sharing on operational and financial performance via innovation performance. Overall, the literature has supported the input-process-output model in the field of KM.

The selection (ranking) process of KM initiatives is a critical step in current KM methodologies [16, 17]. For example, Bornemann and Sammer [17] developed a comprehensive assessment tool to prioritize KM interventions. These were classified into sixteen categories according to four types of scopes (individual, team, organisation and business environment) and four levels (goals, knowledge, business processes and data).

Recently, several MCDM methods have been applied to select KM strategies. Patil and Kant [5] proposed a list of KM initiatives for supply chain firms, ranging from the use of IT systems for knowledge dissemination to establishing a transparent work flow policy. The selection of the initiatives was then performed using fuzzy TOPSIS method. Wu and Lee [18] and Tseng et al. [19] developed methods based on the Analytic Network Process (ANP) to help companies that need to select favourable KM strategies. This method was further combined with the Decision Making Trial and Evaluation Laboratory (DEMATEL) to visualize the structural model behind criteria [20, 21]. Fuzzy VIKOR was used to deal with the vagueness and granularity in the linguistic assessments of KM technologies [22]. A similar approach was proposed by Jenab and Sarfaraz [23] to select the most appropriate set of KM technologies to support the innovation processes in organizations. Hung et al. [24] solved the KM adoption problem by using a model combining several MCDM methods. This model was employed to evaluate the KM gap in various industries. To select an optimal KM investment strategy, Zandi and Tavana [25] developed a hybrid model combining MCDM with fuzzy real option evaluation and group ordinal approach, respectively.

3 Research Methodology

Following Luukka [26] and Collan and Luukka [27], we used fuzzy similarity both to calculate the distance between fuzzy ratings and to make the final evaluation and ranking. In our study, the fuzzy ratings and the weights of criteria were assigned as the values of linguistic variables. To represent these values, we used trapezoidal membership functions.

The selection of KM initiatives is considered to be a group MCDM, which has recently attracted an increasing attention in economic problems (see [28] for an overview). Hereafter, we use the following notation: D_k is the k-th decision-maker, k = 1, 2, …, K; A_i is the i-th initiative (alternative), i = 1, 2, …, m; C_j is the j-the criterion, j = 1, 2, …, n; w_j is the weight of the j-th criterion; and x_{ij} is the rating of the i-th initiative w.r.t. the j-th criterion. Fuzzy positive-ideal solution (FPIS) is defined by the maximum function from the set of normalized positive trapezoidal fuzzy numbers [29]. Similarly, fuzzy negative-ideal solution (FNIS) is defined by the minimum function.

Fuzzy similarity for the final evaluation and ranking was calculated as the average similarity over n criteria. Similarity $S(X, Y)$ between two fuzzy numbers X and Y was determined as follows [26]:

$$S(X,Y) = \left(1 - \frac{\sum_{i=1}^{4}|a_i - b_i|}{4}\right) \times \frac{\min(P(A),P(B))}{\max(P(A),P(B))} \tag{1}$$

where trapezoidal fuzzy numbers X and Y are represented by quadruples (a, b, c, d) and $P(A)$ and P(B) are the perimeters of the fuzzy numbers.

In the first stage of the research methodology, the goal of the MCDM problem is represented for each stakeholder, this is separately for owner, top management, employee and customer. The first stage can be summarized as follows:

1. A committee of decision-makers was formed and the assessment criteria were identified. We used four KM processes as the criteria: knowledge creation, organization, sharing and application.
2. The linguistic variables were chosen for the ratings and weights as follows: very low, low, medium, high and very high.
3. The ratings and weights were aggregated over all decision-makers.
4. Weighted fuzzy decision matrix was constructed.
5. The fuzzy similarity to FPIS and FNIS was calculated and averaged for each KM initiative.

In the second stage, the output fuzzy similarities from step 5 were used as inputs to the third layer of the fuzzy TOPSIS model, where the objectives of individual stakeholders were aggregated as follows:

6. The weights were assigned to stakeholder objectives using the linguistic variables from step 2.
7. The ratings and weights were aggregated over all stakeholders as weighted averaged fuzzy similarities S_{ij}.
8. Aggregated fuzzy similarities were calculated for each KM initiative according to the formula [30]:

$$S^* = \left[\sum_{j=1}^{n} (S_{ij})^2 \right]^{1/2}, \tag{2}$$

9. The ranking was determined according to the values of aggregated fuzzy similarity to FPIS.

4 Case Study – A Medical Supplier Company

4.1 Company Description

The company under study is a medium-sized medical supplier company operating in European, US and Asian markets. The structure of the plant under consideration comprised three divisions, namely manufacturing, commercial and support. The manufacturing division covers production, quality control, material purchasing, etc. The commercial division is responsible for human resources, communication with partners and customers, finance and marketing. The support division provides services in information and technology.

4.2 Questionnaire Survey

In the first step of the proposed methodology a committee of decision-makers was formed. Three managers D_1, D_2 and D_3 of the above mentioned divisions were asked via questionnaire to provide the linguistic assessment (very low - VL, low - L, medium - M, high - H, very high - VH) of the ratings and criteria weights.

Sixteen KM initiatives were proposed as alternatives $A_1, A_2, \ldots, A_{16}$. These covered all categories developed by [17] and included: education plan (A_1); regular meetings (A_2); regular workshops with suppliers (A_3); share knowledge verbally through brainstorming (A_4); easier access to data (A_5); install a file server (A_6); regular reports on external information access (A_7); IT infrastructure development (A_8); employee empowerment (A_9); utilizing team working methods (A_{10}); utilizing customer feedback (A_{11}); continuous improvement process initialization (A_{12}); knowledge vision communication (A_{13}); individual learning objectives' coordination (A_{14}); benchmarking (A_{15}); and report on objectives (A_{16}).

In the first stage (steps 3–5), criteria were represented by KM processes: knowledge creation (C_{11}), organization (C_{12}), sharing (C_{13}) and application (C_{14}). In the second stage (steps 6–7), the criteria were formed from stakeholder objectives: owner (C_{21}), top management (C_{22}), employee (C_{23}) and customer (C_{24}). Again, the selection of stakeholders conformed the KM methodology designed by [17].

4.3 Construction of Fuzzy Decision Matrixes

We used trapezoidal membership functions to represent the values of linguistic variables as shown in Fig. 1.

The decision-makers D_1, D_2 and D_3 assigned the following linguistic assessment to the weights of stakeholder objectives:

W2D1 = {L VH M L},
W2D2 = {H VH L L},
W2D3 = {L H H H},

to the KM processes w.r.t. the stakeholder objectives:

W1D1 = {M M M VH;M L M L;M VL VL M;M VL M H},
W1D2 = {L H M VH;M H VH VH;M L H L;L L H VH},
W1D3 = {L L L H;M H M M;M M M L;L M L L},

and to the ratings of initiatives w.r.t. the KM processes:

FDM1 = {H VL L VL;VL VL L VL;L VL L VL;VL VL VL VL;H VL L VL;VL VL L VL;L VL L VL;L VL VL L;H VL M VL;VL VL L VL;L VL M VL;L VL VL L;H VL M VL;VL VL M VL;M VL H VL;L VL VL M},

FDM2 = {VH VL M L;VL VH M VL;M L VH L;L VL L H;H VL VL M;L VH L VL;M VL VH VL;L VL VL VH;M VL L VL;VL VH VL VL;VL VL H VL;VL VL L M;M L L L;VL VH M VL;VL VL H L;VL VL L VH},

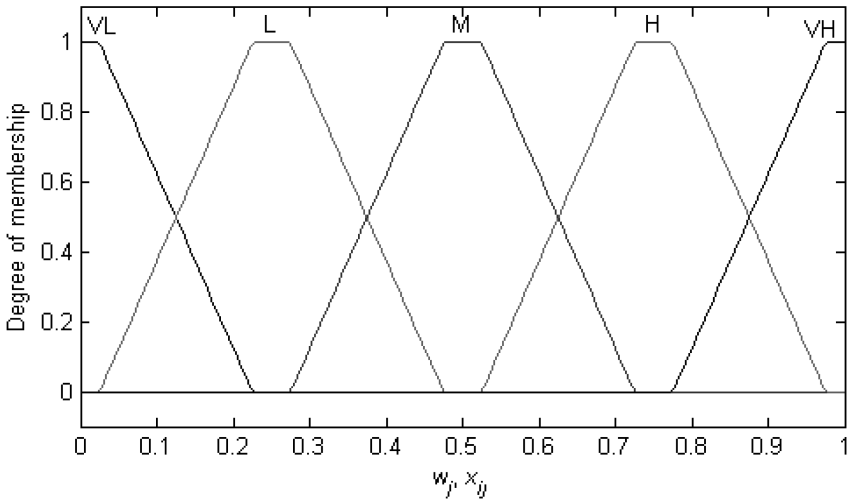


Fig. 1. Linguistic variables and corresponding membership functions for criteria weights and initiatives' ratings.

FDM3 = {M VL VL VL;VL L VL VL;L VL M VL;VL VL VL L;M VL L L;VL M L VL;L VL M VL;L VL VL M;M VL L VL;VL H VL VL;VL VL H VL;VL VL M H;H L L M;VL M L VL;VL VL H M;L VL L VH},

where FDM1, FDM2 and FDM3 denote fuzzy decision matrixes for the three decision-makers.

4.4 Results

First, we examined the effect of each KM initiative on individual stakeholder objectives separately. Table 1 shows that KM initiatives had different impact on individual stakeholders. However, the initiatives A_{13} (knowledge vision communication) and A_{15} (benchmarking) seem to be critical for all stakeholders.

In Table 1, three initiatives with the highest scores are marked in bold. The highest similarity to FPIS was observed for customer objectives (A_9, A_{13} and A_{15}). Employee empowerment (A_9) was relevant for the objectives of owner and customers, in particular. In contrast, regular workshops with suppliers (A_3) were crucial for top management and employees.

Similar results were observed for FNIS (Table 2). Note that a higher similarity to FPIS (i.e. a smaller distance to FPIS) is preferred, while it is a lower similarity for FNIS. In further considerations, FPIS was employed to rank the alternatives (KM initiatives).

Second, we explored the importance of individual stakeholder objectives for the organization (Table 3). The highest weights were observed for the owner and top management.

Table 1. Fuzzy similarities to FPIS of KM initiatives for individual stakeholders.

	C_{21}	C_{22}	C_{23}	C_{24}
A_1	0.722	0.605	**0.717**	0.735
A_2	0.578	0.577	0.633	0.591
A_3	0.620	**0.613**	**0.746**	0.633
A_4	0.536	0.537	0.638	0.549
A_5	0.722	0.593	0.693	0.735
A_6	0.578	0.569	0.617	0.591
A_7	0.620	0.547	0.689	0.633
A_8	0.656	0.531	0.623	0.662
A_9	**0.772**	0.562	0.656	**0.785**
A_{10}	0.578	0.536	0.580	0.591
A_{11}	0.670	0.521	0.633	0.683
A_{12}	0.656	0.605	0.596	0.662
A_{13}	**0.772**	**0.687**	**0.714**	**0.785**
A_{14}	0.628	0.596	0.638	0.642
A_{15}	**0.772**	**0.658**	**0.714**	**0.785**
A_{16}	0.754	0.587	0.660	0.754

Table 2. Fuzzy similarities to FNIS of KM initiatives for individual stakeholders.

	C_{21}	C_{22}	C_{23}	C_{24}
A_1	0.794	0.661	**0.677**	0.794
A_2	0.937	0.716	0.782	0.937
A_3	0.875	**0.592**	**0.646**	0.875
A_4	1.000	0.723	0.749	1.000
A_5	0.794	0.663	0.695	0.794
A_6	0.937	0.707	0.778	0.937
A_7	0.875	0.726	0.734	0.875
A_8	0.831	0.771	0.800	0.839
A_9	**0.746**	0.735	0.740	**0.746**
A_{10}	0.937	0.766	0.837	0.937
A_{11}	0.828	0.777	0.788	0.828
A_{12}	0.831	0.661	0.777	0.839
A_{13}	**0.746**	**0.538**	**0.651**	**0.746**
A_{14}	0.890	0.705	0.779	0.890
A_{15}	**0.746**	**0.608**	0.686	**0.746**
A_{16}	0.761	0.693	0.736	0.775

Table 3. Fuzzy similarities to FPIS for individual stakeholders.

	C_{21}	C_{22}	C_{23}	C_{24}
FPIS	0.901	0.978	0.809	0.846

Finally, the ratings and weights from Tables 1 and 3 were aggregated by Eq. (2) in order to obtain the final ranking according to the aggregated fuzzy similarities to FPIS and FNIS (Table 4). The similarities and ranking show the priority tools in the KM strategy of this company. Communication of the KM objectives and benchmarking are the most critical issues in KM implementation. Furthermore, additional KM tools should be implemented such as employee empowerment (A_9), education plans (A_1) and reporting on objectives (A_{16}). These initiatives support the objectives of the stakeholders at most. Regular workshops with suppliers (A_3) were among three most effective KM initiatives according to the similarity to FNIS. Similarly, sharing knowledge verbally through brainstorming (A_4) was also more important according to FNIS when compared with the FPIS criterion.

Table 4. The overall fuzzy similarities to FPIS and FNIS and ranking of KM initiatives.

	FPIS	Rank	FNIS	Rank
A_1	1.230	4	1.304	3
A_2	1.055	13	1.503	14
A_3	1.155	7	1.342	7
A_4	1.000	16	1.557	16
A_5	1.215	6	1.311	4
A_6	1.044	14	1.498	13
A_7	1.098	11	1.432	9
A_8	1.096	12	1.440	11
A_9	1.236	3	1.320	6
A_{10}	1.013	15	1.546	15
A_{11}	1.111	9	1.432	9
A_{12}	1.122	8	1.378	8
A_{13}	1.315	1	1.193	1
A_{14}	1.111	9	1.451	12
A_{15}	1.301	2	1.237	2
A_{16}	1.225	5	1.316	5

Taken together, our results suggest that initiatives knowledge vision communication and benchmarking were the most important knowledge management initiatives. Knowledge vision represents the starting point for all knowledge management initiatives and provides direction to the organization's knowledge management efforts. It should be communicated by the top management for synchronizing the entire organization and it also connects knowledge management with intellectual capital of the organization [31]. Benchmarking (both internal and external) is, on the other hand, a common application of knowledge management systems supporting the transfer of best practices [32].

5 Conclusion

We addressed the issue of contradictory stakeholder objectives in KM using a hierarchical fuzzy TOPSIS model, which has been successfully applied in several business MCDM problems [33–35]. Moreover, we demonstrated that this model can handle the uncertainty and vagueness associated with the process of KM implementation.

The case study showed the effectiveness of the model in ranking KM initiatives according to organizational objectives. Thus, this decision-support tool may help overcome the barriers which were reported to be the main reason for a low success rate of KM adoption [5]. Taken together, this paper has underlined the importance of assessing KM initiatives with respect to multiple stakeholder objectives. Thus, this research will serve as a base for future studies on KM adoption.

Although our research could be a useful aid for support decision makers because it provides a stakeholder perspective, a number of caveats need to be noted regarding the present study. First, the choice of the criteria can be easily modified to include other KM processes such as knowledge acquisition. Second, previous research has shown that the KM processes may be interconnected [36, 37]. This issue has been addressed by using ANP or DEMATEL. The main challenge is, however, an assessment of these connections. This is an important issue for future research. It is suggested that the association of the KM processes is investigated in future studies using, for example, ANP and DEMATEL in fuzzy context [38]. Finally, we did not consider the costs (neither capital nor operating) of the initiatives which should also be taken into account when selecting an optimum portfolio of knowledge management initiatives. This is also a vital issue for future research.

Acknowledgments. This article was created as a part of the solution of the research task No. 14-02836S, financially supported by the Grant Agency of the Czech Republic.

References

1. Gold, A.H., Arvind Malhotra, A.H.S.: Knowledge management: an organizational capabilities perspective. J. Manage. Inf. Syst. **18**(1), 185–214 (2001)
2. Zack, M., McKeen, J., Singh, S.: Knowledge management and organizational performance: an exploratory analysis. J. Knowl. Manage. **13**(6), 392–409 (2009)
3. Hajkova, V., Hajek, P.: Efficiency of knowledge bases in urban population and economic growth - evidence from European cities. Cities **40**, 11–22 (2014)
4. Dyer, J.H., Hatch, N.: Using supplier networks to learn faster. Sloan Manage. Rev. **45**(3), 57–66 (2006)
5. Patil, S.K., Kant, R.: A fuzzy AHP-TOPSIS framework for ranking the solutions of knowledge management adoption in supply chain to overcome its barriers. Expert Syst. Appl. **41**(2), 679–693 (2014)
6. Ruoning, X., Xiaoyan, Z.: Extensions of the analytic hierarchy process in fuzzy environment. Fuzzy Sets Syst. **52**(3), 251–257 (1992)
7. Chen, C.T.: Extensions of the TOPSIS for group decision-making under fuzzy environment. Fuzzy Sets Syst. **114**(1), 1–9 (2000)

8. Chen, T.Y., Tsao, C.Y.: The interval-valued fuzzy TOPSIS method and experimental analysis. Fuzzy Sets Syst. **159**(11), 1410–1428 (2008)
9. Wang, Y.M., Elhag, T.: Fuzzy TOPSIS method based on alpha level sets with an application to bridge risk assessment. Expert Syst. Appl. **31**(2), 309–319 (2006)
10. Zolfani, S.H., Sedaghat, M., Zavadskas, E.K.: Performance evaluating of rural ICT centers (telecenters), applying fuzzy AHP, SAW-G and TOPSIS grey, a case study in Iran. Technol. Econ. Dev. Econ. **18**(2), 364–387 (2012)
11. Lee, H., Choi, B.: Knowledge management enablers, processes, and organizational performance: an integrative view and empirical examination. J. Manage. Inf. Syst. **20**(1), 179–228 (2003)
12. Hackman, J.R., Morris, C.G.: Group tasks, group interaction process, and group performance effectiveness: a review and proposed integration. In: Berkowitz, L. (ed.) Advances in Experimental Social Psychology, vol. 8, pp. 45–99. Academic Press, New York (1975)
13. Andreeva, T., Kianto, A.: Knowledge processes, knowledge-intensity and innovation: a moderated mediation analysis. J. Knowl. Manage. **15**(6), 1016–1034 (2011)
14. Darroch, J.: Knowledge management, innovation and firm performance. J. Knowl. Manage. **9**(3), 101–115 (2005)
15. Wang, Z., Wang, N.: Knowledge sharing, innovation and firm performance. Expert Syst. Appl. **39**(10), 8899–8908 (2012)
16. Tiwana, A.: The Knowledge Management Toolkit: Practical Techniques for Building a Knowledge Management System. Prentice Hall PTR, Upper Saddle River (2000)
17. Bornemann, M., Sammer, M.: Assessment methodology to prioritize knowledge management related activities to support organizational excellence. Meas. Bus. Excell. **7** (2), 21–28 (2003)
18. Wu, W.W., Lee, Y.T.: Selecting knowledge management strategies by using the analytic network process. Expert Syst. Appl. **32**(3), 841–847 (2007)
19. Tseng, M.L., Wu, W.W., Lee, C.F.: Knowledge management strategies in linguistic preferences. J. Asia Pac. Bus. Innov. Technol. Manage. **1**, 60–73 (2011)
20. Wu, W.W.: Choosing knowledge management strategies by using a combined ANP and DEMATEL approach. Expert Syst. Appl. **35**(3), 828–835 (2008)
21. Wu, W.W.: Segmenting critical factors for successful knowledge management implementation using the fuzzy DEMATEL method. Appl. Soft Comput. **12**(1), 527–535 (2012)
22. Büyüközkan, G., Feyzioglu, O., Cifçi, G.: Fuzzy multi-criteria evaluation of knowledge management tools. Int. J. Comput. Intell. Syst. **4**(2), 184–195 (2011)
23. Jenab, K., Sarfaraz, A.R.: A fuzzy graph-based model for selecting knowledge management tools in innovation processes. Int. J. Enterp. Inf. Syst. **8**(1), 1–16 (2012)
24. Hung, Y.H., Chou, S.C.T., Tzeng, G.H.: Knowledge management adoption and assessment for SMEs by a novel MCDM approach. Decis. Support Syst. **51**(2), 270–291 (2011)
25. Zandi, F., Tavana, M.: A hybrid fuzzy real option analysis and group ordinal approach for knowledge management strategy assessment. Knowl. Manage. Res. Pract. **8**(3), 216–228 (2010)
26. Luukka, P.: Fuzzy similarity in multicriteria decision-making problem applied to supplier evaluation and selection in supply chain management. Adv. Artif. Intell. **2011**(6), 1–9 (2011)
27. Collan, M., Luukka, P.: Evaluating R&D projects as investments by using an overall ranking from four new fuzzy similarity measure based TOPSIS variants. IEEE Trans. Fuzzy Syst. **22** (3), 505–515 (2013)

28. Turskis, Z., Zavadskas, E.K.: Multiple criteria decision making (MCDM) methods in economics: an overview. Technol. Econ. Dev. Econ. **2**, 397–427 (2011)
29. Chen, C.T., Lin, C.T., Huang, S.F.: A fuzzy approach for supplier evaluation and selection in supply hhain management. Int. J. Prod. Econ. **102**(2), 289–301 (2006)
30. Wang, J.W., Cheng, C.H., Huang, K.C.: Fuzzy hierarchical TOPSIS for supplier selection. Appl. Soft Comput. **9**(1), 377–386 (2009)
31. Hussi, T.: Reconfiguring knowledge management-combining intellectual capital, intangible assets and knowledge creation. J. Knowl. Manage. **8**(2), 36–52 (2004)
32. Alavi, M., Leidner, D.E.: Review: knowledge management and knowledge management systems: conceptual foundations and research issues. MIS Q. **25**(1), 107–136 (2001)
33. Bao, Q., Ruan, D., Shen, Y., Hermans, E., Janssens, D.: Improved hierarchical fuzzy TOPSIS for road safety performance evaluation. Knowl. Based Syst. **32**, 84–90 (2012)
34. Kahraman, C., Ates, N.Y., Çevik, S., Gülbay, M., Erdogan, S.A.: Hierarchical fuzzy TOPSIS model for selection among logistics information technologies. J. Enterp. Inf. Manage. **20**(2), 143–168 (2007)
35. Paksoy, T., Pehlivan, N.Y., Kahraman, C.: Organizational strategy development in distribution channel management using fuzzy AHP and hierarchical fuzzy TOPSIS. Expert Syst. Appl. **39**(3), 2822–2841 (2012)
36. Hajek, P., Henriques, R., Hajkova, V.: Visualising components of regional innovation systems using self-organizing maps - evidence from European regions. Technol. Forecast. Soc. Change **84**, 197–214 (2014)
37. Matatkova, K., Stejskal, J.: Descriptive analysis of the regional innovation system-novel method for public administration authorities. Transylvanian Rev. Adm. Sci. **39**, 91–107 (2013)
38. Büyüközkan, G., Çifçi, G.: A novel hybrid MCDM approach based on fuzzy DEMATEL, fuzzy ANP and fuzzy TOPSIS to evaluate green suppliers. Expert Syst. Appl. **39**(3), 3000–3011 (2012)

Successful Knowledge Sharing and Knowledge Management Practices

Heterogeneous Business Process Management: A Metamodel-Based Approach

Antonia Azzini[1], Paolo Ceravolo[1](✉), Ernesto Damiani[1], Nello Scarabottolo[1], Francesco Zavatarelli[1], Claudia Vicari[2], and Vincenzo Savarino[2]

[1] SESAR Lab - Dipartimento di Informatica, Università Degli Studi di Milano, Milano, Italy
{antonia.azzini,paolo.ceravolo,ernesto.damiani,nello.scarabottolo, francesco.zavatarelli}@unimi.it
http://sesar.di.unimi.it

[2] Engineering S.p.a., Palermo, Italy
{claudia.vicari,vincenzo.savarino}@eng.it
http://www.eng.it

Abstract. Dynamically adapting business process to changing needs, and promptly reacting to events are today key factors to maintain competitiveness in the market. Business Process Management (BPM) is focusing today more and more on a BPM in the large approach to process changes that embrace all of the specific techniques and mechanisms needed to design, enact, execute and monitor processes and process-aware information systems. A novel and promising feature of BPM in the large is the ability to store, aggregate and combine this huge and very diverse amount of data that can enable new ways of analysing current operations and can deliver new business insights. The KITE.IT Project [2] is aimed at facing such challenges in the context of the Italian aerospace industry, using and integrating Open Source tools exclusively. The project has recently deployed its initial open framework offering a robust data integration system in an open and scalable architecture. In such a context a metamodel approach was considered the very first base to design a system apt at integrating data originated from heterogeneous sources. The clear advantage that can be reached is the improvement of the speed and the effectiveness of business operations. The clear advantage, which can be reached, is the improvement of the effectiveness of business operations. This paper, in particular, presents the design process that was implemented in defining the KITE.IT Metrics Metamodel (KMM). A final evaluation of the framework, as it was initially deployed, is also reported.

1 Introduction

Business Activity Monitoring (BAM) and Business Process Management (BPM) are terms that refer to the notion of providing real-time access to data about business activity or business process, with specific attention to critical business performance indicators [9,10,16]. The aim is improving the speed and effectiveness of business operations. From a technological point of view, BAM and BPM

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 73–86, 2015.
DOI: 10.1007/978-3-319-21009-4_7

rely on the advance of the Information Technology (IT) to enable the convergence of operational business intelligence [20] and real-time application integration.

The analysis of processes is central to BPM, as it allows the diagnosis of problems and their remediation. In fact, process intelligence has emerged as a new branch of process analysis that is focused on improving and standardizing the inspection of processes to provide deeper insights than traditional performance monitoring and measurement. For instance, a business needs to adjust its supply chain instantaneously if its preferred supplier turns out to be unable to deliver or switch to an auction or a broker mechanism to find a replacement within minutes, because it cannot afford to let its customers wait. The wide adoption of processes and process-aware information systems is drastically and rapidly shifting the traditional view of BPM from clerical business processes (e.g. procurement, sales and post-sales assistance, human resource management) to the so-called BPM in the large. BPM in the large assumes flexible, dynamically configurable processes consuming data coming from heterogeneous sources. In the literature these issues are typically related to real time data warehouse, i.e. the integration of data from different devices, underlining the potential of ubiquitous computing. As more data sources are added and more events are recorded, very detailed information about running processes are being collected and the digital universe is starting to align with the real organization processes. However, today the potential of this integrated view is under-expressed. The reason is that business process analysis tools cannot cope with heterogeneous data generated by process-aware information systems. These tasks are typically executed by different people, reside on different levels of abstraction, and are performed within different information systems; therefore a number of challenges arise, especially with respect to interoperability and practicability.

The KITE.IT Project [2], is aimed at facing such challenges in the context of the Italian aerospace industry, using and integrating Open Source tools exclusively. In such a context the definition of a set of metamodels is a key step towards integrating data originated from heterogeneous models. Following [15], we consider a model as a "description of (part of) a system written in a well-defined language, while a well-defined language is a language with well-defined form (syntax), and meaning (semantics), which is suitable for automated interpretation by a computer". Therefore, on one side, a model must focus on describing parts of a system with the desired level of detail (avoiding superfluous information). On the other side, the meta model has to introduce the foundation for creating models in a meaningful, precise and consistent manner. Therefore, metamodelling provides models with a common structure using an abstract categorisation, and in the same time orients those models with a detail specification support. Our aim is to define a set of open metamodels able to overcome the shortcomings above reported, but that also allow satisfying all the important features that a complete metamodel should cover, like the possibility to join multiple metrics to an objective, allowing then to evaluate the metrics quality, the abstraction capability from a language for the metric execution, and, finally, the visualization of the results [3].

In this work particular attention is given to the requirements elicitation for the KITE.IT Metrics Metamodel (KMM). Our aim was to support four main procedures such as *(i)* creation and computation of metrics, *(ii)* connection between metrics and paradigms of analysis, *(iii)* contextualization and *(iv)* visualization of the results. We first revised the literature verifying the lack of a metamodel covering all such procedures. Our attention on these procedures was motivated by the need of introducing additional levels of flexibility on the monitoring process. These general objectives were then detailed in a requirement elicitation phase, tested via a questionnaire submitted to the BPM practitioners and experts, who were chosen among KITE.IT consortium partners and partners' clients. KITE.IT consortium brings together some of the largest companies in the Italian aeronautical sector.

More specifically, the paper is organized as follow. Section 2 compares five approaches already presented in literature, while Sect. 3 and Sect. 4 present, respectively, the requirements elicitation and the metamodel implementation. Tests performed on the framework are reported in Sect. 5. Final remarks about the performed tests are reported in Sect. 6.

2 Related Works

In this work we aim to underline the capability of our approach to cover at the same time three of the most important functionalities that a meta model implementation should cover, regarding in particular *(i)* the possibility to join multiple metrics to an objective, *(ii)* the contextualization, the abstraction from a language for the metric execution, and, finally, *(iii)* the visualization of the results. For this reason we map them against five approaches reported in the literature, as shown in Table 1.

Table 1. Features comparison of some approaches in literature.

Approach	Creation, contextualization and execution of metrics	Connection between metrics and analysis paradigms	Visualization of the results
Ref [13]	n/a	Abstract constructs from technical details	Graphical notation
Ref [12]	System goals are defined by grouping metrics	n/a	Defines a language to define the meta model
Ref [1]	Novel solutions for metrics integration	n/a	n/a
Ref [8]	Meta model for computing multiple-project metrics	n/a	n/a
Ref [5]	n/a	Abstract language for querying	Graph visualization for results exploration

In the work carried out in [13] the authors presented a high-level graphical modeling language that is specifically tailored to cover the full spectrum of BAM applications and enables an integration with business process models, targeted towards the specification of both process monitoring and control aspects. Based on easily interpretable semantics, it provides connection points to BPMN to allow for a better integration and combination of the models. It also supports the definition of different types of process-related KPIs which can either refer to single process instances or can be located on an aggregated process level. However, the authors also reported that further efforts are needed to demonstrate and systematically evaluate the language, when applied in real-world business scenarios.

The paper [12] analyses existing work on KAOS language definition and consolidates it into a single meta model, defined in Ecore, that is suitable as a basis for MBSE. It also implements a graphical editor. In particular, the goal model is a graph of nodes and links, that describes systems goals, the conflicts between these goals, and the obstacles to the satisfaction of them. Also in this work the authors reported that existing language definitions for KAOS suffered from two key problems: the first regards the degree of formalization, since the language definition is based on an abstract syntax model instead of a concrete one. The second problem regards the incompleteness. The language definition provides an abstract syntax model of those parts of KAOS that are of relevance to the problem addressed by the research article.

In the work carried out in [1] the authors strive for a better integration of Process Performance Indicators (PPIs) into the business process lifecycle (definition, execution, analysis). PPIs should be modeled together with the business processes. Thus, they present an ontology to define PPIs which comprises a comprehensive classification of PPIs and explicitly defines how the PPIs are related to elements of a BPMN business process model. The authors also provide a valuable contribution for closer integration of business process modeling and monitoring aspects. However, while a connection between BPMN constructs and PPIs is established, the sole definition and modeling of process metrics is of limited value in a BAM context as threshold values and the corresponding exception handling mechanisms need to be expressed as well.

The approach presented in [8] provides process measurement in a joint-project, multi-process model business environment. It is based on a simple meta model for computing across-process, multiple-project metrics designed to permit monitoring of CMMI compliance. The open source tool used by the authors has been developed to support their approach and is capable of producing the measurements needed for monitoring of a set of large-scale development projects using different process models, in a real industrial setting in Europe. The results support the view that it will not always be possible to aggregate the same set of metrics across disparate process models.

The interface between the process and the measurement modules is the specification of measurable attributes of process activities and work products, needed to compute the metrics. Finally, all process metrics to be included in process reports are defined in terms of KPIs. This meta model approach is suitable for cross-process measurements, but it is important to establish which metrics and

KPIs (coming from different projects) can be safely aggregated and which ones cannot, due to semantic differences. Explicit support for relationships between KPIs and metrics of different processes is a basic requirement for a platform supporting heterogeneous models.

In [5] the authors presented a framework based on a process event query language and graph-based querying processing engine for the explorative querying and understanding of BP execution from various user perspectives. The proposed query language for analyzing events is called FPSPARQL, which is a Folder-Path enabled extension of the SPARQL query language. FPSPARQL has been used to query and analyze events, folder and path nodes in order to analyze business process execution data. The authors also described the results of the evaluation of the performance of the engine, and the quality of results over large event logs. Finally they provided a front-end tool for the exploration and visualization of results in order to enable users to examine the event relationships and the potential for discovering process instances and process models. The implemented graph visualization tool is aimed at exploring the results. Using this tool, users are able to view folders, paths, and the result of queries in a list and visualized format. A visual query interface to support users in expressing their queries over the conceptual representation of the event log graph in an easy way is still one of the future activities reported by the authors, as well as interactive graph exploration and visualization techniques.

3 BPM Metrics Requirements Definition

3.1 The KITE.IT Project

This Section presents the results of the requirements elicitation analysis for the KMM. In this regard, taking high level objectives, the metamodel should be capable of supporting procedures such as *(i)* creation and computation of metrics, *(ii)* connection between metrics and paradigms of analysis, *(iii)* contextualization and *(iv)* visualization of the results.

3.2 The Questionnaire

The investigation is focused on the identifications of functionalities that are strategically important in the implementation of the procedures above mentioned. All such requirements are defined through a questionnaire aimed at capturing the standpoint of 29 BPM practitioners of the aerospace Italian industry. The practitioners, selected among KITE.IT partners and partners' clients, were junior and senior experts of the industry, with a seniority ranging from at least twelve months to over twenty years of experience. The resulting metamodel is expected to be one of the main open source results of the KITE.it project.

The questionnaire is designed on the Kano model [19]. The Kano model is an efficient tool to understand user expectations and to formulate a product strategy to meet those needs and expectations. The analysis starts from a set of questions testing the requirements considered by the designer. Multiple questions are formulated for testing a single requirement, typically one positively oriented and one negatively oriented. For each question a user can choose among different answers, divided into six categories, corresponding respectively to: `Attractive`, `Performance`, `Must be`, `Indifferent`, `Questionable`, `Reversal`.

where each answer represents a different level of interest on the considered requirement. In particular, `Attractive` indicates that the requirement is not expected, but considered appealing; `Performance` indicates that the level of satisfaction increases when increasing the availability of such a requirement; `Must be` refers to a requirement that is considered mandatory for the product, while `Indifferent` refers to a requirement that is considered not particularly relevant. The latter two answers indicate that an error occurred in filling the questionnaire, since contradictory answers are given for the same requirement. For this reason such two answers are considered not valid from the design point of view, except in case of high occurrence that indicates an error in designing the questionnaire.

In order to define the requirements to be tested in the questionnaire we started from the components implementing the procedures supported by the KITE.IT framework [2]. Figure 1 illustrates eight components interconnected by thirteen relations[1]. From the relations we then derived the thirteen requirements tested in the questionnaire. The literature review presented in Sect. 2 underlined that none of the existing proposals is covering all these procedures. We then focused on the interconnections among the identified components.

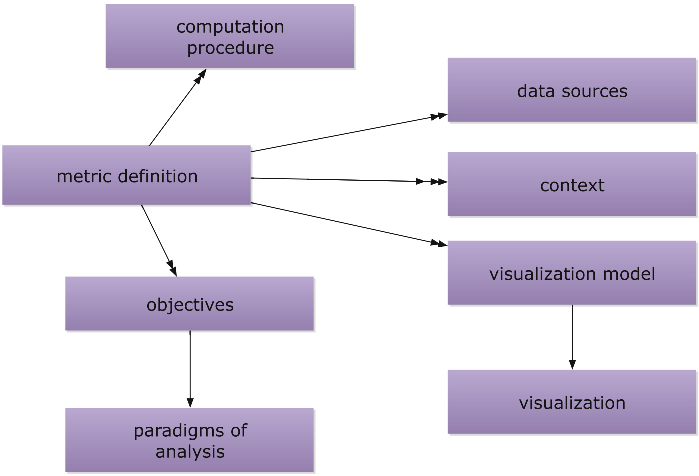


Fig. 1. Component's correlation for the questionnaire definition.

[1] we have a relation for each arrow head in the figure.

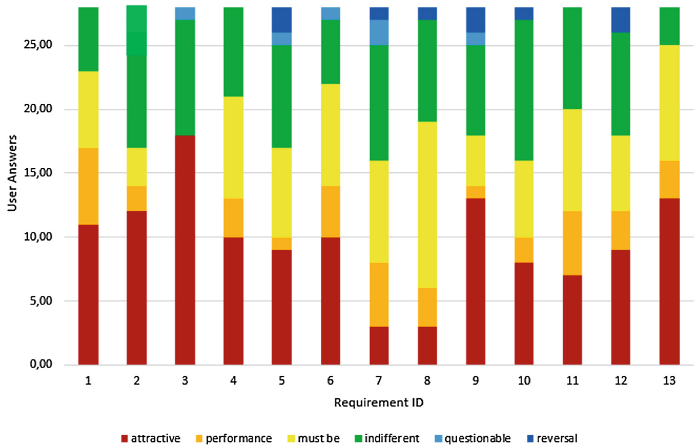


Fig. 2. Categories distribution for each tested requirement.

Table 2 details these thirteen requirements. The last column reports the category with predominant occurrence, expressed in terms of relative percentage, as emerged from the answers we collected.

A more complete overview on the results is provided in Fig. 2, which shows the distribution of the six answer categories for each of the thirteen requirements analyzed. Categories, ordered from `Attractive` to `Reversal`, are mapped on colours ordered from red to blue.

The analysis of the results shows that the collected answers are oriented to assign significance to all the requirements tested. Even at a first sight it can be easily observed that all the requirements are oriented to warm colours[2], that cover a space ranging from 50 % to 80 % of the answers. The `Attractive` category is predominant in eight requirements we tested, with coverage below 20 % only for requirements 7 and 8, where `Must be` is the predominant category. KITE.IT partners considered very positively these results, taking the decision to have all requirements implemented in the KMM.

4 The Metamodel for BPM Metrics

As pointed out in Sect. 1 the KITE.IT metamodel is aimed at supporting four procedures. Introducing this metamodel we aim at providing an abstraction from technical details defining the required interrelations among the different data model operating in the system. From the point of view of the syntactical notation we decided to adopt XML Schema [21]. This choice resulted from the examination of the expressiveness, that was considered appropriate, and due to the possibility of exploiting the schema in the implementation of the user interface

[2] Thus, including `Attractive`, `Performance`, and `Must be`.

Table 2. Requirements tested with our questionnaire and corresponding predominant category.

ID	Requirement Description	Predominant Category
1	Connect metric and paradigm of analysis. *Each objective is associate to one metric in the adopted paradigm* $(GQM, BSC, ..)$	Attractive 39 %
2	Multiple Objectives. *Metrics associated to more than one objective in same analysis*	Attractive 43 %
3	Multiple paradigms of analysis. *Metrics can be associate to objectives belonging to different paradigms of analysis* $(GQM, BSC, ..)$	Attractive 64 % %
4	Connect metric definition and measurement workflow. *For each metric abstract definition and specific measurement workflow given, natural language can be used for metric definition*	Attractive 36 %
5	Independent metric definition and measurement workflow *For each metric abstract definition the measurement workflow can be changed*	Attractive 32 %
6	Connect metric definition and data source. *An abstract definition is given to each metric and associate it with the specific data sources that will be used to evaluate the metric elements*	Attractive 36 %
7	Metric definition and data source are independent.*For each metric abstract definition, specific data sources are used for evaluation*	Indifferent 32 %
8	Connect metric definition and observation context. *For each metric the observation context can be given*	Must be 46 %
9	Connect metric definition and observation contexts. *Series of contexts of observation can be provided over which the metric has to be calculated*	Attractive 46 %
10	Independent metric definition and observation contexts. *Observation context can be modified*	Indifferent 39 %
11	Connect metric definition and result visualization. *A visualization approach can be associate to each metric* $(histogram, graph, etc.)$	Must be 29 %
12	Independent metric definition and results visualization. *The visualization approach can be modified for each metric* $(histogram, graph, etc.)$	Attractive 32 %
13	Connect metric measurement workflow and result visualization. *Given metric measurement workflow and data visualization approach, results visualization can be automatically obtained*	Indifferent 46 %

devoted to the definition of metrics. The KMM XML Schema is presented in Fig. 3 and is published on the KITE.IT web site[3].

In particular, the metamodel has been organized according to four main areas:

- Measurement area: it allows to describe the strategic objectives. As detailed in [2] these objectives can be expressed according to different measurement paradigms, such as for instance Balanced Scorecard [14] or Goal Question Metric [4]. In particular, in KITE.IT, four different perspectives are considered: Financial, Value Network, Processes, and Learning and growth. These objectives can be expressed both in terms of policies supporting the objectives or policies mitigating the risks associated with the objectives identified.
- Observation area: it permits to describe the metric observation conditions in terms of process instances or timeframe of validity. Using this definition it is possible to filter the data to be analyzed but also to define a thresholds of validity for critical metric values, like those indicating the success in an objective achievement.
- Computation area: it allows to describe the operations that have to be carried out over the elements analyzed by the metric. It also describes the data access procedures, in terms of localization of the data sources required for the metric computation. The KITE.IT framework was designed to access data based on the RDF model [17], according to the data integration system designed in [18]. However the metamodel can address alike relational databases [7] or REST services [11].
- Visualization area: it allows to describe the data structures that will be used for visualization, connecting the parameters of the operations defined under the computation area with data structures supported by the KITE.IT framework. Right now two data structures are supported allowing to visualize data in terms of graphs and trees or in terms of histograms [6].

5 Tests

A preliminary version of the KITE.IT framework was recently deployed and a first round of tests was performed on the framework in order to verify whether the initial system met the requirements or not.

A first test was executed in the main academic project partner, and a controlled example composed of data about students' curricula was extracted from the university data system and uploaded to the project site. Data instances are identified by student IDs and include learning experience in recent years, in particular the classes attended by each student with dates, the grades obtained, and the final thesis date with the evaluation.

[3] http://www.kite\discretionary--project.it/documents/10665/0/Kite_Metric_Schema.

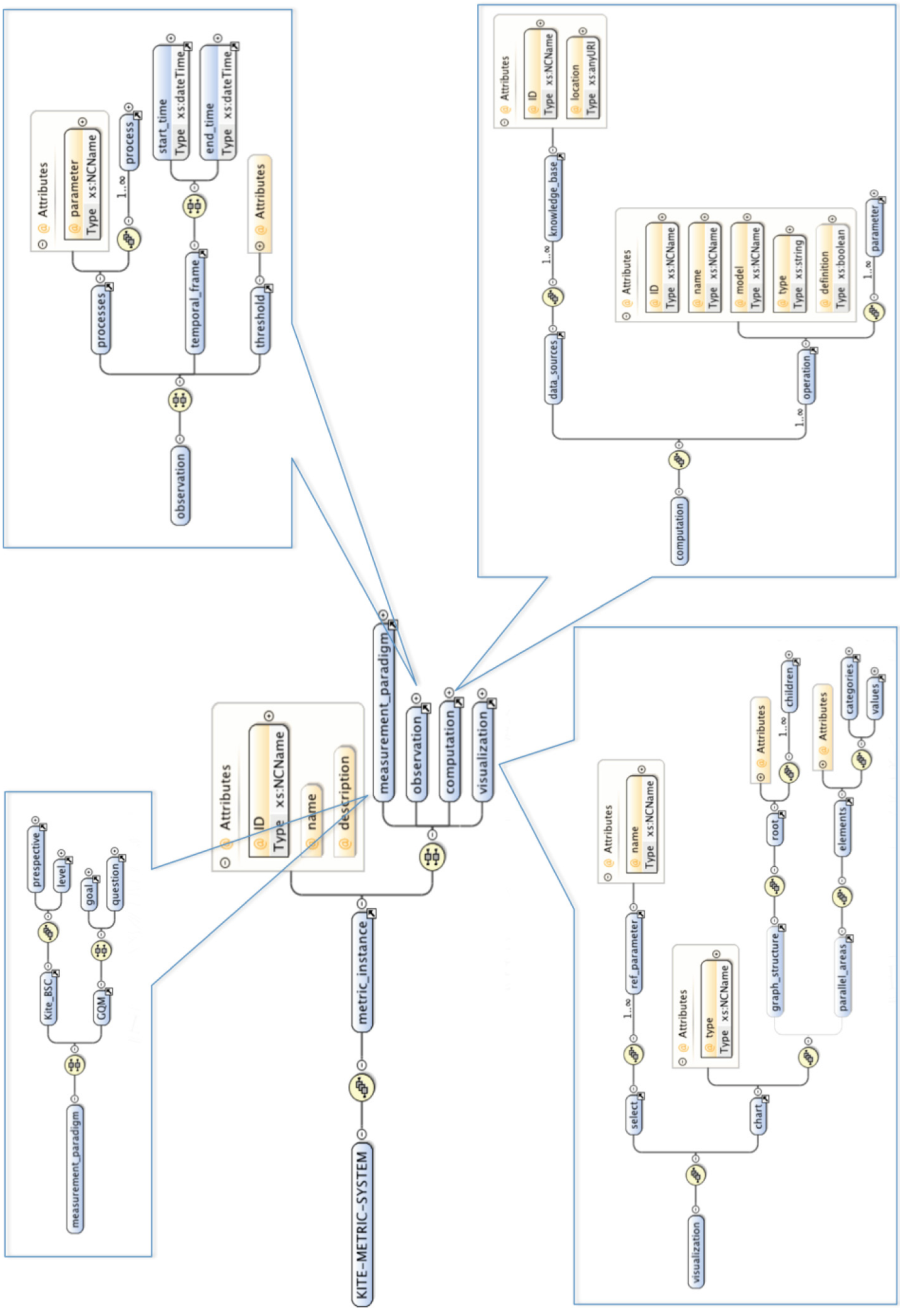


Fig. 3. Major details of the KITE.it metric metamodel (KMM).

The curriculum process was modelled using the integrated KITE.IT Modeller. KITE.IT Modeller allows a lightweight and open standard representation model designed to support real time monitoring of business processes and based on a shared vocabulary called Resource Description Framework (RDF) [18]. The usage of RDF as modelling language allows fast model set-up and data independence. Process data were uploaded to KITE.IT framework with a data set of around 3000 records. Then the 13 requirements have been grouped into 5 test cases, each one related to a specific use case. The first test group included the requirements 1, 2 and 3, all them related to metrics, objectives and paradigm of analysis, and basically concerned the capacity of the system to operate with different paradigms such as Goal Question Metrics or Score Cards. The second test group included the requirements 4 and 5, and was about how the framework allows defining and measuring metrics. The third test group included the requirements 6 and 7, and concerned how to process data sets that have to be provided for metric computation. The fourth test group included the requirements 8, 9 and 10, and the fifth test group the requirements 11, 12 and 13, dealing respectively with metric observation contexts and the final graphical metrics visualisation. The 5 evaluation forms were prepared and given to three BPM senior practitioners. These evaluation forms guided the tester through the system, pointing to the 5 use cases and providing operational instructions, such as how to add a new metric definition and edit all the related attributes. The tester was requested to evaluate the requirements using a qualitative grade scale, indicating at which extent the requirements were met.

The qualitative grade scale that was used in the evaluation phase was the following:
grade 0 = the requirement is not satisfied at all,
grade 1 = the procedure to implement this requirement is in a preliminary stage,
grade 2 = the requirement is only partially satisfied,
grade 3 = the requirement is satisfied, but some components are missing,
grade 4 = the requirement is fully satisfied, but the operational procedure needs to be optimised,
grade 5 = the requirement is fully satisfied.

Table 3 details the tests outcome. The last column of the table reports the evaluation grades in the 0–5 scale. Since the primary objective was to have an initial evaluation of the framework, the outcome indicates that most of the requirements were fully met. However some issues were raised, in particular the requirement n. 3 had a low grade. This requirement requests that a metric could be associated to objectives belonging to different paradigms of analysis, such as Goal Question Metrics and Score Cards, and was indicated as "attractive" by our survey. It is a high level requirement and the score is related to the fact that the methodology used in KITE.IT uses its own and specific paradigm of analysis, called KITE Score Cards, as a default. The issue reveals that, not unsurprisingly, in these complex frameworks there is a strong interdependency between project components, and some high level project choices can interfere with other requirements. All the other KMM requirements were met and the grades were acceptable.

Table 3. Tests outcome

ID	Test N	Requirement	Outcome
1	1	Connect metric and paradigm of analysis	4
2	1	Multiple objective	3
3	1	Multiple paradigms of analysis	1
4	2	Connect metric definition and measurement workflow	5
5	2	Metric definition and measurement workflow are independent	5
6	3	Connect metric definition and data source	5
7	3	Metric definition and data source are independent	5
8	4	Connect metric definition and observation context	5
9	4	Connect metric definition and series of observation contexts	5
10	4	Metric definition and observation contexts are independent	5
11	5	Connect metric definition and result visualisation	5
12	5	Metric definition and results visualisation are independent	5
13	5	Connect metric measurement workflow and result visualisation	4

6 Conclusions

In this work, we defined an metric meta model based on a set of requirements. The rationale behind our work is supporting BPM on heterogeneous models. As mentioned, we are interested in a representation for abstracting from the calculation method but also from the objectives that a metric is measuring and the context of application of this metric. Thus, we identified a number of elements the KMM is implementing, carrying off several advantages with respect to the approaches already presented in the literature.

- KMM is supporting the connection between a metric and multiple objectives. As objective can relate on different paradigm of analysis, this support the integration of different monitoring systems.
- Connecting an objective to multiple metrics KMM can support the assessment of metrics, comparing results from multiple monitoring systems, and supporting the continuous improvement of the monitoring system.
- The metric definition can be more general with respect to other approaches presented in the literature, since it is not only based on the well known KPIs, but also on business rules aimed at verifying the metamodel constraints.
- Metrics can be designed together with the visualization approach that will be used in monitoring the results.

The data acquired from the survey have been used during the implementation of the KITE.IT framework, enhancing the requirements that resulted of most interest. Finally the first version of the open KITE.IT framework was deployed, with the implementation of a full metric lifecycle based on the KMM, and some initial qualitative tests were performed. Future works will test KITE.IT framework using industrial data sets and will address the issues raised in the initial test phase.

Acknowledgement. This work was partly funded by the Italian Ministry of Economic Development under the Industria 2015 contract - KITE.IT project.

References

1. Resinas, M., Ortega, A.D.R., RuizCortes, A.: Defining process performance indicators: an ontological approach. In: Meersman, R., Dillon, T., Herrero, P. (eds.) Confederated International Conferences: CoopIS, IS, DOA and ODBASE, Hersonissos, Crete, Greece, October 25-29, 2010, Proceedings, Part I. Lecture Notes in Computer Science, vol. 6426, pp. 555–572. Springer, Heidelberg (2010)
2. Arigliano, F., Azzini, A., Braghin, C., Caforio, A., Ceravolo, P., Damiani, E., Savarino, V., Vicari, C., Zavatarelli, F.: Knowledge and business intelligence technologies in cross-enterprise environments for italian advanced mechanical industry project presentation. In: Proceedings of the 3rd International Symposium on Data-driven Process Discovery and Analysis, Co-located with 39th International Conference on Very Large Databases (VLDB 2013), SIMPDA 2013, vol. 1027, pp. 104–110. CEUR (2013)
3. Azzini, A., Ceravolo, P., Damiani, E., Zavatarelli, F., Vicari, C., Savarino, V.: Driving knowledge acquisition via metric life-cycle in process intelligence. In: I-KNOW, p. 26 (2014)
4. Basili, V.R.: Software modeling and measurement: the goal/question/metric paradigm (1992)
5. Beheshti, S.-M.-R., Benatallah, B., Motahari-Nezhad, H.R., Sakr, S.: A query language for analyzing business processes execution. In: Rinderle-Ma, S., Toumani, F., Wolf, K. (eds.) BPM 2011. LNCS, vol. 6896, pp. 281–297. Springer, Heidelberg (2011)
6. Chi, E.H.: A taxonomy of visualization techniques using the data state reference model. In: IEEE Symposium on Information Visualization, InfoVis 2000, pp. 69–75. IEEE (2000)
7. Codd, E.F.: A relational model of data for large shared data banks. Commun. ACM **13**(6), 377–387 (1970)
8. Colombo, A., Damiani, E., Frati, F., Oltolina, S., Reed, K., Ruffatti,G.: The use of a meta-model to support multi-project process measurement. In: Proceedings of the 15th IEEE Asia-Pacific Software Engineering Conference, APSEC 2008, pp. 503–510. IEEE (2008)
9. Van der Aalst, W.M.P.: Process Mining: Discovery Conformance and Enhancement of Business Processes. Springer, Heidelberg (2011)
10. Dumas, M., La Rosa, M., Mendling, J., Reijers, H.A.: Fundamentals of Business Process Management. Springer, Heidelberg (2013)

11. Fielding, R.T., Taylor, R.N.: Principled design of the modern web architecture. ACM Trans. Internet Technol. (TOIT) **2**(2), 115–150 (2002)
12. Clark, T., Nwokeji, J.C., Barn, B.S.: Towards a comprehensive meta-model for kaos. In: Proceedings of the International Workshop on Model-Driven Requirements Engineering, MoDRE 2013, pp. 30–39. IEEE (2013)
13. Matzner, M., Friedenstab, J.P., Janiesch, C., Muller, O.: Extending bpmn for business activity monitoring. In: Proceedings of the 45th Hawaii International Conference on System Science, HICSS 2012, pp. 4158–4167. IEEE (2012)
14. Kaplan, R.S., Norton, D.P.: The Balanced Scorecard: Translating Strategy into Action. Harvard Business Press, Watertown (1996)
15. Kleppe, A.G., Warmer, J., Bast, W.: MDA Explained: The Model Driven Architecture: Practice and Promise. Addison-Wesley Longman Publishing Co. Inc., Boston (2003)
16. Laguna, M., Marklund, J.: Business Process Modeling, Simulation and Design, 2nd edn. CRC Press, Boca Raton (2013)
17. Lassila, O., Swick, R.R., et al.: Resource description framework (rdf) model and syntax specification (1998)
18. Leida, M., Majeed, B., Colombo, M., Chu, A.: A lightweight RDF data model for business process analysis. In: Cudre-Mauroux, P., Ceravolo, P., Gašević, D. (eds.) SIMPDA 2012. LNBIP, vol. 162, pp. 1–23. Springer, Heidelberg (2013)
19. Sauerwein, E.: The kano model: how to delight your customers. In: Proceedings of the IX International Working Seminar on Production Economic, pp. 313–327 (1996)
20. Surajit, C., Umeshwar, D., Vivek, N.: An overview of business intelligence technology. Commun. ACM **54**(8), 88–98 (2011)
21. Thompson, H.S.: Xml Schema Part 1: Structures Second Edition (2004)

Application of Knowledge Management for Increasing the Effectiveness of Franchise Based Organizations

Waleed Alneaimi(✉)

The British University in Dubai, Dubai, UAE
2014132164@student.buid.ac.ae

Abstract. The research investigates how the application of knowledge management and knowledge management initiatives is beneficial for franchises. How can an integrated model of knowledge Management be helpful for franchisor-franchisee-customer triad? The methodology applied is quantitative and franchises with KM and with KM were assessed to evaluate the effectiveness of KM for knowledge retention. The group of franchises with KM were further investigated to assess how the use of ICT, the role of information shared by Franchisee and the role of information shared by franchisor is helpful in better and effective management of franchise based organizations.

1 Introduction

A Knowledge Management System can be defined as an information system that adds value to the overall system by maintaining and developing processes that facilitate formation, storage, recovery, dissemination, and application of knowledge both inside and outside the organization (Quaddus and Xu 2005; Alavi and Leidner 2001). The inclusion of knowledge management into the domain of strategic management has been progressing and now it considered as the next step in a sequence of societal developments (Wiig 1997). In the present era that has been defined as "Knowledge Economy" knowledge management and knowledge management systems have emerged as a fundamental and imperative tool for management (Ndlela and Du Toit 2001). The management of knowledge is also important because the rate of knowledge creation has accelerated and is anticipated to experiencing even more rapid accelerations (Bontis 2002).

Knowledge Management has become increasing popular throughout the globe and many developed, developing and emerging nations have adopted and inoculated knowledge management in the core strategies of their business. Knowledge management has also become a hot-topic and buzz-word for management professionals throughout the Middle East. Many sectors including oil and gas have adopted various forms of knowledge management systems to increase the system efficiencies.

1.1 Background of the Study

The franchising business model is widely used and popular model that has been successfully being practiced throughout the globe. Franchising business model is also

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 87–101, 2015.
DOI: 10.1007/978-3-319-21009-4_8

different from other forms of business models as efficacious franchisors don't just trade products and services. Franchisors create a business system and improve it till perfection and then sell the rights and benefits along with the tacit knowledge of their perfect business system to prospective franchisors. When one converses about a franchise business model, it important to consider the sharing of knowledge between the franchisor and the franchisee.

There have been many researches on franchise business model effectiveness but most of them have focused on the legal and contractual relation or the rights of ownership rather than the management of knowledge. Knowledge is a complex asset and has the ability to serve as a competitive business advantage. Better knowledge management can help franchisers and franchisees become more efficient and better serve their customers and develop a sustainable competitive edge (Morgana and Huntb 1999). It is therefore considered that knowledge management must be applied with franchise management systems to increase their efficiencies.

1.2 Significance of the Study

The Middle East and GCC in particular are important markets for franchising (MENAFA 2015). The opportunities for food franchising, retail franchising, service franchising, B2B Franchising and B2C Franchising are all experiencing growth whereas their counterparts in the west and even Asia are experience a downtime due to slow economic growth (Brad 2012). Franchising in the Middle East is considered the best option for business growth and transnational expansion by the International Franchise Association (IFA 2007). The GDP of GCC was estimated to be US $1.7 trillion in 2014 (IIF 2014) and the franchise market of the Middle Eat was estimated to have grown by 27 % or US$30 billion more than previous year (Arabian Business Review 2014). Many international brands have already moved to the MENA and GCC market and there is more developments anticipated in the franchising sector (Jones 2003; Mellahia et al. 2011). The positive and favorable outlook is estimated to continue, hence it is considered important to study how franchising model can benefit from better knowledge management.

1.3 Aim and Objectives

The aim of this study is to investigate how the application of knowledge management and knowledge management initiatives is beneficial for franchises. How can an integrated model of knowledge Management be helpful for franchisor-franchisee-customer triad?

The study also aims to find how tacit management is being controlled and how can the management of tacit knowledge improved in a franchising business model. The study also aims at exploring some of the barriers and critical factors necessary for the implementation of an effective knowledge management system across franchising business model relevant specifically to the Middle East to increase and augment its efficiency.

1.4 Research Question

The following are the key questions that this research aims to answer;

- What are the key differences between a traditional franchise and a KM based franchise
- What are the critical factors for implementation of KM in a traditional franchised business model
- Can networked franchise with a Knowledge Management system really increase the efficiency and effective that is measurable as well.

1.5 Hypothesis

There are 6 hypothesis for the study;

1. H1: There is significant loss of crucial and critical information in the traditional franchise model
2. H2: In a networked franchise with KM, the exchange of knowledge (both tacit and explicit) are likely to increase
3. H3: The role of ICT is significant and it facilitates the management and sharing of knowledge
4. H4: The information shared by the franchisor is significant and helpful for making strategies and strategic planning
5. H5: Information shared by the franchisee is significant and helpful for the franchisor
6. H6: Better knowledge management is concurrent with better results and better service.

1.6 Hypothetical Framework

The hypothesis are divided into 2 sets, one that compares the efficiency of knowledge management in franchisees where no knowledge management system is available and compares it with organizations/franchises that have KM implemented. The comparison of knowledge management in both types of organization can help know about the difference in knowledge leveraging and which type of franchises are able to better exploit knowledge (Fig. 1).

Secondly, we have also wish to evaluate the role of ICT. The role and information shared by the franchisee, the role and importance of the information shared by the franchisor on the overall effectiveness of knowledge management. This is done through analyzing the relationship and association by regression analysis to determine whether there is causal relationship of not (Fig. 2).

The equation clearly indicates that in the hypothetical framework developed, Knowledge Management systems is the Dependent variable and the use of ICT, information shared by franchisee and franchisor are the independent variables (Fig. 3).

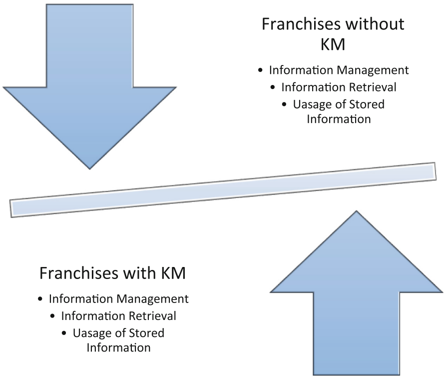


Fig. 1. Comparison of Franchises with and without KM

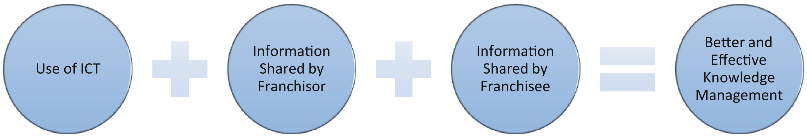


Fig. 2. Conceptual model of components of effective KM in Franchise businesses

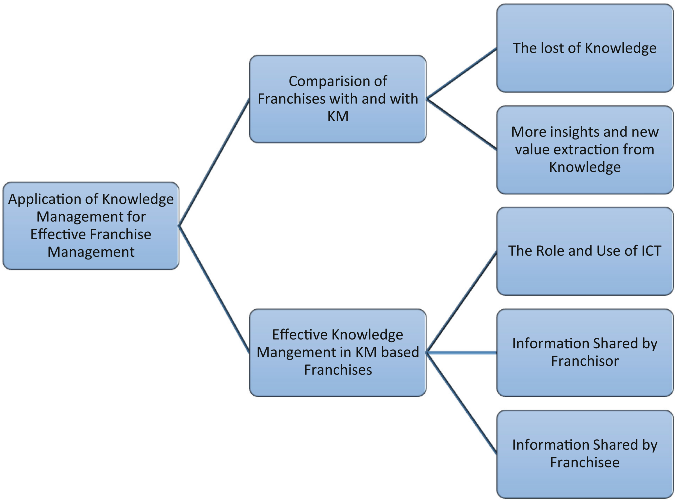


Fig. 3. Illustration of hypotheses this research is testing for validation

2 Literature Review

2.1 Franchising Business Model and the Sharing of Knowledge

Ilan Alon defines franchising as a hybrid manifestation of a market and a firm (Alon 2010) as it depicts key characteristics of both a market for exchange of goods and firms and a business enterprise. Justis and Judd (2007) consider franchising as an opportunity for 2 parties, the owner of product, service or any copyrights protected entity and a local distributor. The franchisor or the owner of the business to get proceeds and profits and also expand the business and the franchisee or distributor can leverage upon the business system of the owner (Judd and Justis 2007).

The aspect of knowledge sharing is very important in a franchising business model (Chen and Hammerstein 2002). It is not only the details of the copy-righted product or service that is being granted but also the knowledge about the industry, customers, suppliers, workforce, growth strategy, changing in demographics, curving curves and other variables are shared. O'Dell and Grayson (1998) believe that when a franchisor is granting the rights, he is basically sharing the secret of his/her success and hence complete information and knowledge about the business must be shared with the franchisee (O'Dell and Grayson 1998).

It is common for franchisers and franchisees to use information systems for sharing and dissemination of information (Seideman 1998). The sharing of knowledge, the importance of knowledge management and also the risks associated with sharing of too much information in a franchise model have been discussed in literature (Nault and Dexter 1994) but much of the discussion is focused on the operational aspect or single versus multiple franchisees.

2.2 Traditional, Business Format and Networked Franchising

Franchising have been considered an important business model and was established business model back in the late nineteenth century (Lafontaine 2008). It was the singer sewing company that recognized the importance of this model and adopted it widely, resulting in a widespread popularity of franchising since 1851.

In the traditional or the common type of franchising, the owner of a business grants a license to a local deal to trade a particular service or product while using the owners name and other brand identities. The local dealer of the franchisee gives a certain royalty or fees based upon their contractual agreement to the owner, usually products that are pre-made like automobiles follow this model. Another type of franchising is the business format franchising where the owner of the business shares the details about the processes and operating methods along with the permission to use elements of brand identity. The contractor or the franchisor provides a complete business model and the franchisee along with the support replicates it. The franchisee can in return gives royalty and some fees on the sales of the goods or services (Preble and Hoffman 1995). Third type of franchising includes network franchises. Where the owner grants multiple dealers either exclusive or mutual rights of distribution, these rights can be defined per territory or a specific number of its product and services. All franchisees communicate back with the franchisor who keeps on churning the information received from various franchisees to improve the business model (Quinn 1999). The franchising model have evolved with the passage of time and now franchising in done globally and is a multibillion dollar industry and everything from food, sports, restaurants, malls, retail, products, education, media and other services have widely exercised this model (Ozanne and Hunt 2011) and many countries keep on evolving their national models whereas others adopt of imitate them (Welch 1989).

2.3 Knowledge Management in Franchising

In modern era when franchising has emerged as a mega-business, many scholars have studied how the sharing of knowledge among franchisee and franchisor is done, what factors can increase or decreases tacitness of knowledge being shared and the role of information systems in knowledge sharing (Windsperger and Gorovaia 2010). In the present era that has been coined as the knowledge generation and the time for knowledge economies, knowledge is considered as a key asset. Murray and Myers (1999) did a comprehensive survey and found that majority of business organizations consider knowledge management as important source of competitive advantage. 89 % of the participating firms believed that knowledge management to be a source competitive advantage and 85 % believed that a quantifiable value be attached to knowledge Management.

Franchisers around the world leverage upon the information and knowledge generated by other franchisees (Windsperger and Gorovaia 2008).

Knowledge Management has emerged as an important element in modern times, knowledge is considered as a strategic resource in the present knowledge economy (Anon 2004)

Access to information at the right time is precarious and vital, it can help the managers make the right decision at the appointed time and expeditiously, especially when there is a lot of unneeded data or information overload (Quast 2012).

It has been established through research that franchises that have Knowledge Management integrated into their strategy are better at understanding the complex and changing needs of their customers and can reach in a more timely and appropriate manner (Fuller 2012). From McDonalds to Body Shop and from Seven-Eleven to General Motors, businesses that applied KM and inculcated the process of capturing knowledge are better at making decisions that impact the fate of their organization in terms of bottom-line (Despres and Chauvel 2000). It is established that the use of KM in franchise based organizations creates value to the business model and helps organization enhance their knowledge curve (Weaven et al. 2014).

3 Research Methodology

A researcher when investigating a research has a variety of options when selecting the research methodology. The appropriate methodology depends upon the research questions or the aims of the research. The aim of this research is to investigate how the application of knowledge management and knowledge management initiatives is beneficial for franchises. How can an integrated model of knowledge Management be helpful for franchisor-franchisee-customer triad? Hence the methodology selected is quantitative research as it would quantify the results help in reaching a conclusive culmination.

The appropriate research methodology plays a vital role in the research process and influences the results of the research as well. Development of an appropriate research methodology is crucial in enhancing the effectiveness of business research. Research methodology includes the process of the type of research philosophy used, the type research approach, and the research methods used for data collection. Through determination of appropriate research philosophy and approaches that are relevant to the establishment of research objectives is pivotal for research methodology. Following the selection of research philosophy and the approaches to carry out research, it becomes important to choose a suitable data collection method to maintain the quality of research. The choice of research methodology for this study was determined by setting up research questions and the type of data required finding the answer of these research questions.

3.1 Quantitative Approach

The quantitative approach of the data collection includes the survey from the people of the organizations about the information system. The Likert Scale Questionnaire instrument used for collection of the data purposes (Jackson 2007).

3.2 Data Collection and Research Instrument

The data collection instrument used was a self-administered questionnaire. 2 questionnaires were developed, one that was used for a comparison of franchises that didn't use any type of knowledge management and the franchises that were actively engaged in the application of knowledge management.

The second questionnaire was only for the franchises that were using KM and the questionnaire that a Likert scale for responses. On the scale of 1–5, where 1 meant strong disagreement and 5 meant strong agreement, the use and effectiveness of the overall system was requested to be judged. The dimension that were under focus included the use and role of ICT in better knowledge management, the role and importance of two way communication between the franchisee and the franchisor that is the importance of information being shared by both franchisee and the franchisor.

3.3 Data Analysis Tools and Techniques

A number of statistical tools and analysis techniques have been used in the study. Some of the key techniques used included;

1. Comparison of Means using independent t-test
2. Reliability Analysis using Cronbach's Alpha
3. Correlation Analysis
4. Regression Analysis

The computer software SPSS was used for data analysis.

4 Findings and Analysis

4.1 Comparison Between Franchises

The first comparison was done between 50 franchises that had KM, most of these franchises were transnational or multinational franchises belonging to various sectors including oil and gas (petrol stations), retail outlets, food and restaurants, clothing and apparel, luxury and fashion, news agency and media outlets. The other group consisted of 50 franchises that didn't have an accredited and specialized knowledge system. These included mostly local retails shops, food outlets, equipment sellers and hardware distributors.

The results indicate that franchises with KM (Yes group) had a mean of 3.82 which is indicating a mostly favorable response and franchises without KM (No Group) had a mean or average score of 2.3 which was indicating a more disagreement in terms of knowledge retention. The mean scores also indicate that there is a huge difference in the knowledge retention and even the level of significance indicate is .000 which represents a strong significance of the results (Table 1).

The results of the correlation also indicate a negative but strong association of 95.5 % among the two groups.

Table 1. Correlation of franchises with and without KM on grounds on knowledge retention

		Knowledge retention	Knowledge management status
Knowledge retention	Pearson Correlation	1	−.955**
	Sig. (2-tailed)		.000
	N	100	100
Knowledge management status	Pearson correlation	−.955**	1
	Sig. (2-tailed)	.000	
	N	100	100

4.2 Reliability Analysis

When we conserve about reliability we actually mean the consistency of an instrument. How reliable and dependable the instrument is and will the same instrument produce similar results if tested again (Gaur and Gaur 2009). To test the reliability of our research instrument, the statistical measure of Cronbach alpha was used (Table 2).

Table 2. Overall reliability of the research instrument

Reliability statistics	
Cronbach's alpha	N of items
.962	21

The overall Cronbach alpha value generated was 0.962 which shows that the questionnaire had very high level of consistency and is a reliable instrument to gage the KM effectiveness in franchise systems.

The reliability score for the effectiveness of KM based on 5 items was found to be 0.83, that is also a high and reliable score (Table 3).

The reliability scores for the role and importance of ICT based on 4 items was calculated to be 0.737, this is acceptable but a little less than overall score and KM effectiveness scores (Table 4).

Table 3. Reliability for effectiveness of KM

Reliability statistics	
Cronbach's alpha	N of items
.832	5

Table 4. Reliability for role and importance of ICT

Reliability statistics	
Cronbach's alpha	N of items
.737	4

The reliability scores for the role of information shared by the franchisor was found to be 0.826, this again indicates that based on 4 items, the assessment is consistent and reliable (Table 5).

Table 5. Reliability for information shared by the Franchisor

Reliability statistics	
Cronbach's alpha	N of items
.826	4

The last reliability score to be estimated was of the role of information shared by the franchisee back to the franchisor, based on 4 items, it has a score of 0.682. This acceptable in terms of reliability but not very healthy (Table 6).

Table 6. Reliability for information shared by the Franchisee

Reliability statistics	
Cronbach's alpha	N of items
.682	4

4.3 Correlation Analysis

The correlation analysis suggests how are the variables associated with one another, does and change or manipulation in one variable also impact another variable of not. (Black 2009). When we processed the 4 variables that is knowledge management effectiveness, the role of ICT, the role of information shared by the franchisor and the role of information shared by the franchisee. We found a strong and positive correlations among the variables (Table 7).

One very important thing to note in the table is that all levels of significance calculated are .000 which is less than the p-value of 0.005, this indicates that all measures or correlations are statistically significant.

Table 7. Correlation analysis of knowledge management and ICT, Franchisor and Franchisee

		Knowledge management factor	ICT factor	Franchisor factor	Franchisee factor
Knowledge management factor	Pearson correlation	1	.944**	.788**	.895**
	Sig. (2-tailed)		0.000	0.000	0.000
	N	50	50	50	50
ICT factor	Pearson correlation	.944**	1	.870**	.883**
	Sig. (2-tailed)	0.000		0.000	0.000
	N	50	50	50	50
Franchisor factor	Pearson correlation	.788**	.870**	1	.816**
	Sig. (2-tailed)	0.000	0.000		0.000
	N	50	50	50	50
Franchisee factor	Pearson correlation	.895**	.883**	.816**	1
	Sig. (2-tailed)	0.000	0.000	0.000	
	N	50	50	50	50

The reliability "r" for Role of ICT was 94.4 %, for Role and Importance of Information shared by the franchisor was calculated to be 87 % and the Role and importance of Information shared by the Franchisee was estimated to be 81.6 %. All the relationship are positive means that any change in any of these variables will have a concurrent and direct impact of knowledge management system effectiveness.

4.4 Regression Analysis

- From the correlation table we can extract some critical information, here we can say that the correlation among the variables is 0.958 OR 95.8 % (Table 8).
- This result reveal that if there is any modification or alteration in the independent variables, the overall equation of the dependent variable will experience a change of 91.2 %.
- The value of R-Square was appraised to 0.918 this states that 91.8 % of variation can be explained on the basis of independent variable.
- The adjusted R-square is an indication of accuracy and reliability of the data (Table 9).

Table 8. Regression model summary/fitness of model assessment

Model	R	R square	Adjusted R square	Std. error of the estimate
1	.958[a]	.918	.913	.16684

[a]Dependent variable: knowledge management factor

The regression test checks the relationship of dependent and the independent variables (Gaur and Gaur 2009). Here in our model, the dependent model was Effectiveness of Knowledge Management Systems in franchised based organizations and the

Table 9. Regression analysis

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	0.292	0.178		1.639	0.108
	ICT Factor	0.787	0.101	0.835	7.751	0
	Franchisor Factor	-0.162	0.071	-0.199	-2.28	0.027
	Franchisee Factor	0.326	0.094	0.32	3.482	0.001

independent variables were the role and importance of ICT, information shred by the franchisor and the information shared back by the franchisee.

The regression equation is:

$$``\mathbf{Y} = \mathbf{a} + \mathbf{b_1X_1} + \mathbf{b_2X_2} + \mathbf{b_3X_3} + \mathbf{b_4X_4}"$$

Where Y represents the model we have developed to gauge the effectiveness of KM or the Dependent Variable, "a" is the constant that will hold true in all conditions and even when all independent variables are 0 or absent. The alphabet "b_n" is an indication of the per unit change and X_N represents the independent variables.

We already know from the summary of the regression model that the validity of this model is high at 95.8 % and the Beta or constant is 0.292, means that even if all independent variable have null influence, the model will remain true to 29.2 %. Independent variable of Role of ICT have an influence of 78 %, role of information

Table 10. Summary of hypotheses

Hypothesis	P-value & significance	Hypothesis status
H3: The role of ICT is significant and it facilitates the management and sharing of knowledge	(p-value 0.000) <0.05	Acknowledge and accept the hypothesis
H4: The information shared by the franchisor is significant and helpful for making strategies and strategic planning	(p-value 0.027) <0.05	Acknowledge and accept the hypothesis
H5: Information shared by the franchisee is significant and helpful for the franchisor	(p-value 0.001) <0.05	Acknowledge and accept the hypothesis
H6: Better knowledge management is concurrent with better results and better service	(p-value 0.027) <0.05	Acknowledge and accept the hypothesis

shared by franchisor has an impact of negative 16 % and the role of information shared by the franchisee has a positive impact of 32.6 %.

On the basis of this model, we can conclude our remaining 4 hypothesis (Table 10).

5 Discussion and Conclusion

A Knowledge Management System can be defined as an information system that adds value to the overall system by maintaining and developing processes that facilitate formation, storage, recovery, dissemination, and application of knowledge both inside and outside the organization (Quaddus and Xu 2005; Alavi and Leidner 2001). The study aimed to investigate the effectiveness and impact of KM in franchised organizations. There were 6 hypothesis and 5 of them were accepted.

- H1: There is significant loss of crucial and critical information in the traditional franchise model
- H2: In a networked franchise with KM, the exchange of knowledge (both tacit and explicit) are likely to increase
- H3: The role of ICT is significant and it facilitates the management and sharing of knowledge
- H4: The information shared by the franchisor is significant and helpful for making strategies and strategic planning
- H5: Information shared by the franchisee is significant and helpful for the franchisor
- H6: Better knowledge management is concurrent with better results and better service.

This research was done solely in UAE and as UAE is known for its extreme diversity, it was anticipated that the results would differ. UAE is a part of Gulf Cooperation Council (GCC), a geographical union of Arab states in Persian Gulf except Iraq. It is believed that doing business in GCC is different than doing business in West and Asia due to the cultural differences. Even the language is written from right to left and has a complete different set of alphabets, the traditions, culture and religion differ from the western world (Colombo 2014; IBP 2007).

Startlingly all of the hypothesis were accepted, it was assumed on the basis of prevailing perceptions that the nomenclature of doing business in UAE is different from Western countries and even it varies from other Arab and Middle Eastern companies hence that the fourth hypothesis and fifth hypothesis would have little importance and would be turned down. The results are in uniformity with few other researches done on the subject and the fact that KM can help franchises even in UAE to increase their effectiveness.

The application of KM can be really helpful for franchise businesses, as franchises are legally and contractually bound to not make certain decisions without the consent of the franchisor, an integrated KM could help both the franchisee and the franchisor better understand the needs and the changing preferences of key audience in any geographical region and act accordingly. The application of KM can also help in identification of key trends like time periods of high requirements (Christmas or Eid or New Year or National Day etc.) where demand for the product or service is high or vice versa, identify seasons when demand is low, they can further investigate the factor contributing to low or high demand and improve. The advantage that franchise based organizations have is that they can learn from other franchises' experience as well, if a KM is in place, various franchises won't be committing the same error and learn from trial and error but learn from the experience of other franchise.

It is strongly recommended for franchise based organization to invest in KM and extract new knowledge from the existing data, KM is imperative for franchises big and small for survival in this era or cut-throat competition. KM would help make better decisions, enhance their resilience and increase the overall effectiveness.

6 Limitations and Recommendations

6.1 Limitations

This study was aimed at finding the effectiveness of KM in franchised organizations, the researcher was bonded with constraints of time as well as limitation of the sample under study, the for the sake of this study 50 franchises were selected on the basis of a simple random sample based on the convenience. Hence the results are an estimation and the results can't be applicable when considering specific sectors and franchising models. The study ignored the dependent and independent variables of knowledge management in franchises that didn't have an ICT based accredited and formal KM and only compared the level of knowledge retention and level of knowledge loss. There could have been more analysis and comparisons between local and foreign franchises, types of KM being used, the frequency of knowledge generation and compared with the value creation and a comparisons of various sectors within the franchising business model. A food chain will ensure the same taste but will need to adjust as per the tastes of local market like McDonald need to serve Halal meet in UAE and vegetable patties in India.

6.2 Recommendations

The limitations of this study are also the recommendations for future study. A comparison of a local franchise chain versus a foreign franchise chain and the similarities and differences can be investigated. Another dimension that ca be explored in future the various products of the franchises and how KM can help for example an automaker's franchise is in UAE but the manufacturing plant or the franchiser resides in the USA or JAPAN, the impact of such models where franchisee is selling the complete product and a restaurant or an amusement park that follows complete business format from seating to flavors and characters/mascots. The dimensions that can be investigated are limitless and it is inferred that Knowledge Management is here to stay and it has the power to transform businesses.

References

Alon, I.: Franchising Globally: Innovation, Learning and Imitation. Palgrave Macmillan, New York (2010)

Liebl, F.: Knowledge management for strategic marketing. In: Spotts, H. E. (eds.) Proceedings of the 2004 Academy of Marketing Science (AMS) Annual Conference on Assessing the Different Roles of Marketing Theory and Practice in the Jaws of Economic Uncertainty, pp. 48–57. Springer International Publishing (2015)

Arabian Business Review: The MENA Franchising Opportunity and Success Factors (2014). http://www.arabbusinessreview.com/en/article/mena-franchising-opportunity-and-success-factors

Black, K.: Business Statistics: Contemporary Decision Making. Wiley, Hoboken (2009)

Bontis, N.: Rising star of the chief knowledge officer. Ivey Bus. J. **66**(4), 20–25 (2002)

Brad, D.: Small U.S. franchises head to the Middle East (2012). http://www.bloomberg.com/bw/articles/2012-06-21/small-u-dot-s-dot-franchises-head-to-the-middle-east

Chen, Y., Hammerstein, S.: Knowledge, Learning, and Capabilities in Franchise Organizations. Academic Press, Athens (2002)

Colombo, S.: Bridging the Gulf: EU-GCC Relations at a Crossroads. Edizioni Nuova Cultura, Roma (2014)

Despres, C., Chauvel, D.: Knowledge Horizons: The Present and the Promise of Knowledge Management. Routledge, Woburn (2000)

Fuller, S.: Knowledge Management Foundations. Routledge, Woburn (2012)

Gaur, A.S., Gaur, S.S.: Statistical Methods for Practice and Research: A Guide to Data Analysis Using SPSS. Sage, New Delhi (2009)

IBP: Doing Business and Investing in United Arab Emirates Guide. USA International Business Publications, Washington, D.C. (2007)

IFA: Franchising in the Middle East, Dubai and beyond (2007). http://www.franchise.org/franchising-in-the-middle-east-dubai-and-beyond

IIF: GCC: strong diversified growth (2014). http://www.difc.ae/sites/default/files/GCC%20Report%20Main%20Consolidated%20File_final.pdf

Jackson, S.: Research Methods: A Modular Approach. Cengage Learning, Stamford (2007)

Jones, G.: Middle East expansion – the case of Debenhams. Int. J. Retail Distrib. Manage. **31**(7), 359–364 (2003)

Judd, R.J., Justis, R.T.: Franchising: An Entrepreneur's Guide. CENGAGE Learning Custom Publishing (2007)

Lafontaine, F., Blair, R.D.: The evolution of franchising and Franchise contracts: evidence from the United States. Entrepreneurial Bus. Law J. **3**, 381–434 (2008)

Alavi, M., Leidner, D.: Knowledge management and knowledge management systems: conceptual foundations and research. MIS Q. **25**(1), 107–136 (2001)

Quast, L.: Why knowledge management is important to the success of your company. http://www.forbes.com/sites/lisaquast/2012/08/20/why-knowledge-management-is-important-to-the-success-of-your-company/. Accessed 28 March 2015

Mellahia, K., Demirbagb, M., Riddlec, Liesl: Multinationals in the Middle East: challenges and opportunities. J. World Bus. Middle East Spec. Issue Sect. **46**(4), 406–410 (2011)

MENAFA: Franchise Opportunities in The Middle East and North Africa (2015). http://www.menafa.org/

Morgana, R.M., Huntb, S.: Relationship-based competitive advantage: the role of relationship marketing in marketing strategy. J. Bus. Res. **46**(3), 281–290 (1999)

Nault, B.R., Dexter, A.S.: Adoption, transfers and incentives in a franchise network with positive externalities. Mark. Sci. **13**(4), 412–423 (1994)

Ndlela, L., Du Toit, A.: Establishing a knowledge management programme for competitive advantage in an enterprise. Int. J. Inf. Manage. **21**, 151–165 (2001)

O'Dell, C.S., Grayson, C.J.: If only we knew what we know: identification and transfer of internal best practices. Calif. Manage. Rev. **40**(3), 154–174 (1998)

Ozanne, U.B., Hunt, S.D.: The Economic Effects of Franchising. Marketing Classics Press, Decatur (2011)

Preble, J.F., Hoffman, R.C.: Franchising systems around the globe: a status report. J. Small Bus. Manage. **33**(2), 80–88 (1995)

Quaddus, M., Xu, J.: Adoption and diffusion of knowledge management systems: field studies of factors and variables. Knowl. Based Syst. **18**(2–3), 107–115 (2005)

Quinn, B.: Control and support in an international franchise network. Int. Mark. Rev. **16**(4/5), 345–362 (1999)

Seideman, T.: Tying it all together. Franchising World **30**(4), 8–14 (1998)

Weaven, S., Grace, D., Dant, R., Brown, J.R.: Value creation through knowledge management in franchising: a multi-level conceptual framework. J. Serv. Mark. **28**(2), 97–104 (2014)

Welch, L.S.: Diffusion of Franchise system use in international operations. Int. Mark. Rev. **6**, 7–19 (1989)

Wiig, K.M.: Knowledge management: where did it come from and where will it go? Expert Syst. Appl. **13**(1), 1–14 (1997)

Windsperger, J., Gorovaia, N.: The knowledge transfer strategy of franchising firms: evidence from the Austrian Franchise sector (2008), Accessed 25 March 2015

Windsperger, J., Gorovaia, N.: The use of knowledge transfer mechanisms in franchising. Knowl. Proc. Manage. **17**(1), 12–21 (2010)

Client-Vendor Knowledge Transfer Mechanisms in the Context of Information Systems Outsourcing

Hanlie Smuts[1(✉)] and Paula Kotzé[2]

[1] Mobile Telephone Networks, Johannesburg, South Africa
hanliesmuts@mweb.co.za

[2] CSIR Meraka Institute and Nelson Mandela Metropolitan University, Port Elizabeth, South Africa
Paula.kotze@meraka.org.za

Abstract. High levels of shared knowledge can positively influence outsourcing performance and the combination of processes designed to transfer explicit and tacit knowledge, has the most influence on the level of shared knowledge. Some organisations are unaware of the volume and value of knowledge it receives from various inter-organisational partnerships. Without appropriate strategies to facilitate the acquisition and assimilation of knowledge received at different points of contact between the client and vendor, the organisation may be unable to derive maximum benefits from such partnerships. However, little research has been done up to date on how organisations deal with managing knowledge in outsourcing situations. In an attempt to quantify knowledge transfer mechanisms, this paper focuses on the management of knowledge in a client-vendor knowledge transfer context. By considering the ways in which knowledge can be created and exploited for inter-organisational client-vendor situations, benefit may be realised across the whole IS outsourcing lifecycle.

Keywords: Knowledge transfer · Client-vendor knowledge · IS outsourcing

1 Introduction

The information systems (IS) industry has seen the growth of IS outsourcing as a widely accepted business tool from a one-vendor-one-client arrangement, where the vendor provides all IS services to its client, to complex arrangements involving multiple vendors and multiple clients [1–3]. Rao et al. [4] reason that conflicting arguments for and against outsourcing still exists. Advocates of the practice argue that it results in significant cost reduction with increased management control, effective use of staff, capacity on demand and access to advanced facilities. Opponents, on the other hand, contend that IS outsourcing involves major risks with loss of control, loss of qualified IS staff, loss of flexibility and loss of competitive advantage in information management that may result in overdependence on the outsourcing vendor [5–7].

Vilvovsky [5] undertook an analysis of academic writings on IS outsourcing in public organisations and collated issues and concerns in this regard. It was found that uncertainty around budgets poses a serious obstacle for complex outsourcing projects, as

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 102–119, 2015.
DOI: 10.1007/978-3-319-21009-4_9

both the client and the vendor are required to make significant investments over the first year, and this impact on longer-term planning. Loss of control over the technology and project status to a vendor is one of the most frequently mentioned and most harmful pitfalls, as it may result in overdependence on the outsourcing vendor. This, in turn, leads to vulnerability to possible opportunism such as giving other clients higher priority, imposing excessive fees on anything not stated explicitly in the contract, failing to perform necessary maintenance of systems and even loss of data, or the access to data. Failure to keep enough technical expertise and knowledge about the IS outsourcing project in-house is agreed to as being the main reason for the loss of control [8].

Furthermore, in order to capitalise on the benefits of IS outsourcing, many organisations are moving IS work offshore, but often with unsatisfactory results [9]. Key operational challenges experienced by Indian offshore organisations, for example, include competition for labour in the IS industry, the limited availability of human resources at the appropriate experience level, maintaining and improving quality, lack of effective communication in the offshore context and concerns around data privacy with outsourcing to India or other countries [10]. On the other hand, Fabriek et al. [11] maintain that the increased diversity contributed by an offshore team can lead to a stronger team and better quality. Nevertheless, owing to poor communication it often leads to poor decision quality, poor productivity and poor relationships.

According to Blumenberg et al. [12], there is an indication that high levels of shared knowledge positively influence outsourcing performance and that the combination of processes designed to transfer explicit and tacit knowledge has the most influence on the level of shared knowledge [13–15]. Currie and Pouloudi [16] argue that knowledge transfer in IS outsourcing provides an opportunity to encourage researchers, and managers, to consider the value of knowledge-based assets, and to evaluate the extent to which knowledge can be acquired or lost through IS outsourcing. Aydin et al. [13] and Blumenburg et al. [12] concur that, although the importance of knowledge management in IS outsourcing is highlighted by scholars, little research is being done on how organisations deal with managing knowledge in outsourcing situations.

To address this gap in research, the objective of this paper is to investigate how organisations manage knowledge in IS outsourcing activities, with an emphasis on client-vendor knowledge transfer. In order to analyse and describe the client-vendor knowledge transfer considerations relevant in an IS outsourcing arrangement, a research study was conducted in a telecommunication company in South Africa. Section 2 provides the background to the study, Sect. 3 describes the method followed in conducting the research, Sect. 4 reflects on the findings of the study, whilst Sect. 5 concludes the paper.

2 Background

Outsourcing is defined as the action of transferring organisational work to an outsource vendor [17]. The scope of work outsourced and the delivery of the outsource vendor against the scope is managed by an outsourcing arrangement that stipulates the contract conditions, the required service levels and the required deliverable quality of the arrangement [18]. A special type of outsourcing is IS outsourcing, defined by Sparrow [19: 1] as "the practice of handing over planning, management and operation of certain functions to an independent third party, under the terms of a formalised service level agreement". Sparrow [19] maintains that outsourcing should be seen as a strategic management tool. As such, it should be evaluated in the context of the strategic position of the organisation [17, 20].

However, there is evidence to suggest that organisations are not achieving the desired benefits from outsourcing. InformationWeek [21] surveyed business technology decision-makers at North American companies and established that 19 % of respondents achieved better quality at lower cost in 2012 as opposed to 35 % in 2011. One-third of respondents brought work back onshore, with poor-quality work cited as the reason (43 %), with poor communication with offshore staff a significant second (22 %), and an additional 45 % citing communication issues as a very substantial contributor to their decisions. HfS Research and KPMG [22] conducted a global survey of 1079 organisations in which 95 % of respondents indicated that they expect significantly lower operating costs and 94 % highlighted that they expected better standardised processes through IT outsourcing. Only 55 % of respondents were satisfied with achieving significantly lower operating costs and 53 % achieved better standardised processes.

E-commerce and the Internet demand new skills for the knowledge-driven economy and this requirement has complicated IS outsourcing further [23]. Currie et al. [16] suggest that a more detailed understanding of the full implications of the relationship between IS outsourcing and knowledge-based assets is required than one which simply focuses on quantifiable costs. Beyah et al. [15] deduced that knowledge management constructs are helpful in understanding the reasons why some organisations experience successful IS outsourcing arrangements. They argue that knowledge management constructs provide a sound basis for examining and mitigating some of the risks related to IS outsourcing. They encouraged other researchers to consider utilising knowledge management for understanding other IS outsourcing issues.

In Sect. 2.1 an overview of the IS outsourcing lifecycle is presented, Sect. 2.2 considers knowledge management processes in IS outsourcing Sect. 2.3 factors impacting on knowledge transfer and acquisition in the context of IS outsourcing, and Sect. 2.4 the integration of knowledge transfer processes.

2.1 IS Outsourcing Lifecycle

An IS outsourcing lifecycle guides an organisation to realise the full value that outsourcing can provide, to become an informed purchaser, to plan and design the commercial arrangement, to carefully select the best value for money supplier, and to put in the appropriate management skills and effort [24, 25]. Cullen et al. [24] identified three distinct phases of outsourcing in the IS outsourcing lifecycle: (1) the *architect* phase, (2) the *engage* phase, and (3) the *govern* phase. The first two phases deal specifically with the decision to outsource and the preparation for the outsourcing arrangement, and comprises of the activities required to make the arrangement work. The third phase, govern, addresses the management of the outsource arrangement.

The approval of the business case for outsourcing denotes the end of the architect phase, the scope of work definition, up to the point where the outsource vendor is chosen, ends the engage phase, and the ongoing management of the outsourcing arrangement refers to the govern phase of IS outsourcing. The end of the IS outsourcing contract, either through early termination or by reaching the natural end of its term, represents the completion of the lifecycle of the current outsourcing initiative and the start of the next.

According to Currie et al. [16], organisations need to work more closely with their suppliers when negotiating outsourcing contracts. The reasons for this collaboration are to select the right sourcing strategy for the specific market, business, and technical environments, and to establish how benefits, costs and risks of outsourcing can be evaluated. Currie et al. [16] maintain that the ways in which knowledge can be created and exploited for inter-organisational situations, must be considered at the point of contract negotiations and across the entire IS outsourcing lifecycle.

2.2 Knowledge Management Processes in IS Outsourcing

The delivery of business value is essentially a set of knowledge-based activities in the context of IS, as it involves the integration and harmonisation of knowledge "from many individuals of different disciplines and backgrounds, with varied experiences and expectations, located in different parts of the organisation" [12: 344].

There are two types of interaction patterns when an organisation and an outsource vendor work together: (1) within-team and (2) between-team. *Within*-team interaction refers to the interaction between team members of the organisation, or team members of the outsource vendor, among themselves. This implies and requires a close partnership consisting of both formal processes and informal working relationships between the functional areas in the organisation and IS. These interactions between functional areas in the business and IS are of the utmost importance for cross-functional knowledge transfer, although the impact on IS outsourcing performance has not been researched comprehensively [12, 29]. *Between*-team interaction refers to the differences in the belief systems, values and perceptions of deadlines of different [30].

Therefore, the team structure and communication tools used play an important part in idea generation and client-vendor knowledge integration [30]. Knowledge in this context exists in two primary forms: (1) explicit knowledge, as discussed in Sect. 2.2.1, and (2) tacit knowledge, as discussed in Sect. 2.2.2.

2.2.1 Transfer of Explicit Knowledge

Explicit knowledge has been articulated, codified, and stored in certain media, making it communicable in formal and systematic language [31–33].

The following key explicit knowledge factors applies for the interpretation and integration of knowledge between the organisation and the IS outsource vendor [12]:

- *Organisational maturity* indicates the degree to which organisational processes are systemised and formalised through rules, procedures and management practices. It relates to greater efficiency and the reduction of ambiguity because of the application of rules and standard procedures.

- *Continuous improvement of outsourcing contracts and* service level agreement (*SLAs)* points to the regular application of a continuous improvement process applied to formalise, update and continuously improve the contractual and SLA relationship with the outsource vendor.
- *Standards and definition of key technical terms* refer to industry standards, for example the Information Technology Infrastructure Library (ITIL) and IS service process reference tools for enhancing shared knowledge to create a common language between the organisation and the outsource vendor. Standards in this instance also include communication standards and efficiency.
- *Education* concerns the educational level required of employees, referring to articulated knowledge in their field and will depend on their job description. Academic education, organisation-specific training and on-the-job training are considered to transfer explicit knowledge in order to reach the required level of education. This, in turn, assist to improve the shared domain knowledge of the outsource vendor, for example industry-specific knowledge like telecommunications or the financial industry.
- *Explicit definition of communication partners* includes the clear definition of communication associates between which contents have to be transferred, taking cognisance of management levels and explicit interaction structures. These structures are unambiguous elements that form part of the transfer processes for explicit knowledge.

These key factors in the transfer of explicit knowledge have two dimensions: (1) a *content* dimension, dealing with interpretation, sharing and understanding, and (2) a *sender-receiver* dimension, focusing on the interaction structure between the organisation and the outsource vendor [12].

2.2.2 Transfer of Tacit Knowledge

Tacit knowledge, acquired through experience, is far less tangible than explicit knowledge and it is difficult to communicate and formalise [33]. The articulation of tacit mental models is a key factor in creating new knowledge [34]. Face-to-face engagement is important for the transfer of tacit knowledge.

Key factors identified for client-vendor tacit knowledge transfer are [12]:

- *Joint work* involves the transfer of staff by the client organisation to the outsource vendor at the initiation of the outsourcing arrangement. These transferred employees work in close collaboration with the outsource vendor employees to provide a service to the client organisation. This factor has proved to be an efficient means for transferring tacit knowledge through the exchange of information and advice.
- *Trust* is one of the most important aspects of tacit knowledge transfer. It is influenced by the strong involvement of both parties in the relationship and by close collaboration on a daily basis [35, 36]. A trusting relationship is fostered through mixed workgroups, with no functional segregation, as well as a strong personal bond and common understanding [15, 37]. Other aspects that contribute to maintaining a trusting relationship are communication, cultural understanding, capabilities of the outsource vendor, contract conformance and quality [35].

Some organisations are unaware of the volume and value of knowledge it receives from various inter-organisational partnerships. Without appropriate strategies to facilitate the acquisition and assimilation of knowledge received at different points of contact between the client and vendor, the organisation may be unable to derive maximum benefits from such partnerships [2].

2.3 Factors Impacting on Knowledge Transfer and Acquisition in the Context of IS Outsourcing

Knowledge acquisition and transfer in the IS outsourcing context refer to the transfer, acquisition and application of knowledge from the outsource vendor to the organisation in order to assist the organisation in developing the skills and competencies required to sustain their competitive advantage.

The success of this transfer process depends on four sets of factors, namely, (1) knowledge factors, (2) client factors, (3) vendor factors, and (4) relationship factors [38], as listed in.

Two *knowledge factors* affect the acquisition and transfer of knowledge, namely, the nature of knowledge and the mechanism of knowledge transfer. Three aspects influence *client factors*, namely, absorptive capacity, organisational culture, and motivation and rewards. *Vendor factors* consist of two components, specifically vendor capability and vendor credibility. The last component, *relationship factors*, points to cultural distance, quality of communications and the use of collaborative technologies (Table 1).

2.4 Integration of Knowledge Transfer Processes

Knowledge integration refers to the process of blending skills, know-how and expertise in an organisation with external knowledge sources [2, 30]. Knowledge integration is considered an important precursor for the successful development of an outsourced IS development project. In designing an effective IS outsourcing relationship, knowledge sharing and transfer processes must be integrated into the routines and processes of the organisation [12, 13, 39]. Aydin et al. [13] propose a basic three step process to illustrate such integration:

1. Resolve the responsibilities of each party with the outsource vendor, specifically where they are knowledge related.
2. Determine the knowledge gap by assessing the difference between the required tacit, implicit and explicit knowledge and the current knowledge.
3. Establish whether to close the knowledge gap and how this will be achieved.

The outsourcing arrangement defines the roles and responsibilities of all parties involved. Several factors, such as different time zones, cultural differences and potentially different locations, affect the communication and knowledge exchange. In view of the differences between the arrangement prior to IS outsourcing and the current outsourced situation, the organisation is forced to formally agree with the vendor what knowledge is required, how it will be shared and who will facilitate it [13, 16].

Table 1. Knowledge acquisition and transfer factors in an IS outsourcing context [38]

Factor	Sub-factor	Description
Knowledge factors	Nature of knowledge	The number of interdependent routines, individuals, technologies and resources linked to a particular knowledge
	Transfer mechanism	The degree to which the knowledge can be codified and articulated, as it can be transferred more easily than non-modifiable knowledge
Client factors	Absorptive capacity	The ability of the knowledge recipient to recognise the value, and assimilate and commercially apply the value of the new, external knowledge provided by the knowledge source
	Organisational culture	The values, practices and assumptions that affects the learning atmosphere in the organisation and motivates employees to share and capture knowledge in order to facilitate innovation
	Motivation and rewards	The specific incentive schemes that encourage employees to acquire, utilise and share new knowledge with others
Vendor factors	Vendor capability	The wealth of knowledge and experience of the vendor
	Vendor credibility	The extent to which the organisation regards the supplier as trustworthy, reputable and expert
Relationship factors	Cultural distance	The importance of cross-cultural issues in inter-organisational collaboration.
	Communications quality	The excellence of the interaction among business partners
	Use of collaborative technologies	The investment made in such technologies to foster knowledge acquisition and transfer, as well as effective communication

A knowledge gap transpires in instances where tacit knowledge cannot be articulated and when the knowledge owners leave the organisation before their knowledge is made explicit. In order to close/reduce this knowledge gap, the following approaches may be considered:

- Identify and clarify the key knowledge requirements, as well as the level at which they are required [13].
- Externalising the knowledge gaps in the form of application documentation, and by explicitly defining system and business relations [40, 41].
- Identify people with key knowledge and implement specific plans to transfer their knowledge [42].
- Acquiring new knowledge in critical knowledge areas [13].

In the study presented in this paper, research participants shared approaches to close the knowledge gap between client-vendor.

3 Client-Vendor Knowledge Transfer Exploration

The study presented in this paper investigates how organisations manage knowledge in IS outsourcing activities, with an emphasis on client-vendor knowledge transfer. The interpretive case study research methodology was chosen, as the study attempted to learn from the current situation in real life [43, 44], with the results expressed using descriptive statements [43]. Yin [45] defines five components of research design that are important for case studies namely the questions, the propositions, the unit(s) of analysis, the logic linking the data to the propositions and the criteria for interpreting the findings. The unit of analysis in this single case study is the organisation, with the objective to extract the knowledge assets in an IS outsourcing arrangement.

The study was conducted at an organisation that operates in the competitive telecommunications market in South Africa. Technology evolution and enablement plays a key role within this organisation that operates with an IS multi-sourcing approach, which implies that a significant amount of IS projects and IS operational functions are outsourced.

The replacement of a legacy customer management and retail billing system was outsourced to a systems integrator through a procurement process. The scope of work outsourced included business process modelling, high level solution design and architecture, an implementation programme work plan and a data migration strategy. The average project team size based in South Africa was 110 people, with 90 more resources being located in five different countries around the world. One of the requirements was that the final solution should interface to 32 legacy applications and 22 million customer records, as well as 108 million product instances, which were ultimately migrated. During the life of the programme, 653 change requests were logged, 27 cycles of system integration and user acceptance testing was performed, 34 major software releases were promoted from development to testing environments, and more than 15000 test cases were executed successfully in the final cycles. The programme oversaw the training of almost 1000 geographically distributed people on 12 different courses. The progamme was concluded over a 35-month period.

Within the IS outsourcing arrangement the project team members were (1) transferring knowledge to the outsource vendor, (2) obtaining knowledge from the outsource vendor, and (3) creating new, joint knowledge artefacts as an output of the IS outsourcing arrangement. It was believed that the project team could highlight considerations for client-vendor knowledge transfer relevant to all three IS outsourcing phases (architect, engage, govern) as discussed in Sect. 2.1.

In order to identify client-vendor knowledge transfer mechanisms pertaining to the IS outsourcing phases in so as close the knowledge gap described in Sect. 2.4, a questionnaire was designed, with the following 5 sections:

1. Multiple choice questions in order to establish the home organisation of the respondents and the respondents contribution to the outsource phases.
2. Multiple choice questions to enquire about what knowledge management system research participants used to obtain information for the outsource arrangement across all IS outsourcing phases.

3. Determining the methods of knowledge sharing in the IS outsource arrangement phases through multiple choice questions.
4. Open ended questions to determine the knowledge sharing processes and mechanisms that should facilitate better knowledge transfer in the architect, engage and govern phases of IS outsourcing;
5. Identification of general comments regarding sharing of knowledge for IS outsourcing using an open ended question.

The questionnaire was distributed via using an on-line survey tool. The link to the survey was sent to 62 participants representing multiple roles in both the client organisation and in the outsource vendor organisation. The questionnaire was completed by 42 respondents, yielding a 67 % response rate. Of the 42 participants that completed the questionnaire, 22 were permanent employees of the client organisation, 12 were contractors or consultants contracted to the client organisation and 8 were employed by the outsource supplier. The respondents could contribute to more than one phase of the outsourcing: 20 respondents contributed to the architect phase, 11 to the engage phase and 15 to the govern phase of the IS outsourcing arrangement.

The responses of the multiple choice questions were tallied and analysed for the knowledge sources. The answers to the open-ended questions were collated, relevant parts of the data were identified and common themes were classified through a two-step process: (1) use of descriptive codes to attribute a theme to a segment of text [46], and (2) open coding in order to establish themes from the questionnaire data [47, 48]. Each theme was then evaluated for contribution according to the knowledge acquisition and transfer in an IS outsourcing context factors described in Sect. 2.3.

4 Findings and Discussion

Mechanisms of knowledge transfer referenced by the outsource vendor respondents differ to those used by the client organisation research participants. Section 4.1 reports the client-vendor knowledge transfer mechanisms for the architect phase of IS outsourcing, Sect. 4.2 presents the client-vendor knowledge transfer mechanisms for the engage phase of IS outsourcing and Sect. 4.3 for the govern phase of IS outsourcing. For all three phases of IS outsourcing the mechanisms is reported in the knowledge transfer and acquisition factors as described in Sect. 2.3.

4.1 Client-Vendor Knowledge Transfer Mechanisms for IS Outsourcing Architect Phase

The *architect* phase of IS outsourcing focuses on all aspects up to considering the business case for outsourcing and taking the decision to implement an IS outsourcing arrangement. The success of client-vendor knowledge transfer methods are dependent on knowledge, client, vendor and relationship factors as summarised in Table 2.

Knowledge factors for the *client* organisation focus mainly on understanding and making explicit the capability in the client organisation, by conducting learning sessions ensuring a good understanding of IS outsourcing from subject matter experts, and

Table 2. Vendor-client knowledge transfer mechanisms for the IS outsourcing architect phase

Architect	Client organisation	Vendor organisation
Knowledge factors	Create a comprehensive internal resource skills database in order to address skills gaps with external vendor skills Create a knowledge management repository Define knowledge transfer as a key objective in the business case, with specific reference to post implementation adoption, support and maintenance of the solution. Formalised communication and appropriate collaboration tools and utilities Conduct learning sessions with subject matter experts	Conducting formal knowledge transfer workshops Ensure availability of documented organisation knowledge Utilise client organisation expert knowledge formally through consulting Obtain access to client organisation knowledge and business rule repository
Client factors	Define accurate performance measures and benchmarks Plan the work with appropriate stakeholders Prepare a cost benefit analysis Set clearly defined objectives Quantify the business case for outsourcing Understand exactly what the scope of outsourcing requirement is. Obtain activity based costing information Complete proper due diligence of the functions to be outsourced Establish if vendor has leading subject knowledge and experience	Obtain a better understanding of the goals and objectives across all different departments, a clear understanding of the end to end client business and how different departments integrate with each other to achieve the intended goals and objectives
Vendor factors	Refer to previous experience, lessons learnt and reference models of similar activities Research shortlisted vendors extensively Conduct a proof of concept Obtain intelligence about outsource partners and outsource models, as well as the pros and cons of each	Obtain buy in for the outsourcing and adapt the outsource framework where necessary

(*Continued*)

Table 2. (*Continued*)

Architect	Client organisation	Vendor organisation
Relationship factors	Create a single point of ownership that is responsible for communication of all decisions throughout the process Clearly define project roles and responsibilities in terms of expectation Involve key resources from the beginning	Arrange live demonstrations and presentations of current client organisation business Ensure client organisation resources are available Foster culture of knowledge sharing Define key individuals and work with them in their respective domains Establish joint execution focus Enable direct interaction with client outsourcing team leaders Enable interaction among client business, IS, outsource vendor and finance to pro-actively manage potential downstream issues

by ensuring that organisational data is reflected in explicit knowledge repositories. The main objective of *knowledge* factors for the *vendor* organisation is to understand the client organisation and to ensure access to organisational knowledge repositories and experts.

Client factors for the *client* organisation point to all elements of the business case for outsourcing, such as due diligence, cost benefit analysis, outsourcing scope and clear objectives. On the other hand, *client* factors for the *vendor* organisation focus on a clear understanding of the end to end client business and how different departments integrate with each other in order to achieve the intended goals and objectives.

Vendor factors for the *client* organisation include obtaining capability and performance information regarding the outsource vendor through considering a proof of concept, while the *vendor's* effort is on customising an outsourcing framework for the client organisation.

Relationship factors for the *client* organisation point to clearly defining roles and responsibilities for the arrangement and considering which key resources to be engaged at the initiation. *Relationship* factors for the *vendor* organisation focus on the establishment of a joint execution focus and engagement with the client organisation, fostering a culture of knowledge sharing.

4.2 Client-Vendor Knowledge Transfer Mechanisms for IS Outsourcing Engage Phase

The *engage* phase of IS outsourcing includes activities required to prepare for the IS outsource arrangement. The success of client-vendor knowledge transfer methods depend on knowledge, client, vendor and relationship factors, as depicted in Table 3.

Table 3. Vendor-client knowledge transfer methods for the IS outsourcing engage phase

Engage	Client organisation	Vendor organization
Knowledge factors	Define transparent knowledge transfer channels where client organisation stakeholders are kept up to date on the processes used and rationale for vendor selection and subsequent contract terms Unpack the key objective of knowledge transfer into measurable services and products to be included in the criteria for vendor selection and contract negotiation Negotiate and detail knowledge management requirements already in the early stages of the engagement	Schedule proper focus group meetings where all key stakeholders are mandated to attend and participate in sharing the outsourcing vision
Client factors	Access outsource reference models and best practice models Prepare integrated joint work plan Create clear statement of work and preferred vendor list Create a request for proposal Ensure a very clear outsource model scope and client organisation negotiation position Establish a formal process of evaluation to establish suitability and best fit of an outsource partner Create a small senior and empowered cross-functional sourcing team to review outsource proposals (tender submissions) Define clear hand-over/exit procedure which needs to be clearly documented in the scoping of the outsource project/agreement	Define service level agreement measure clearly Ensure detailed definition of business operations in terms of processes, policies and procedures
Vendor factors	Understand industry norms and trends in order to provide better insight into sourcing arrangements Establish vendor track record based on previous projects Obtain comprehensive knowledge of vendor capabilities and experiences to assist in the establishment of contracts and governance roles within the relationships Conduct information sharing sessions with current customers of potential outsource vendor	

(*Continued*)

Table 3. (*Continued*)

Engage	Client organisation	Vendor organization
Relationship factors	Maintain shared information base and keep up to date at all times	Manage client organisation expectations of outsource arrangement Establish iterative detailed updates to all members of the team, as the strategy changes or negotiation items become apparent that change the course of the outsourcing project

The main objective of *knowledge* factors in the *client* organisation is to, already in the early stages of the engagement, negotiate and detail knowledge management requirements, while the *vendor* organisation focuses on engaging with the client organisation in order to clearly understand the outsourcing objectives. In terms of *client* factors for the *client* organisation the definition of a clear statement of work and joint execution plan for the outsource vendor is a priority. The definition of measureable SLAs and a detailed definition of business operations in terms of processes, policies and procedures are the main emphasis of *client* factors for the *vendor* organisation. *Vendor* factors for the *client* organisation include obtaining a comprehensive knowledge of vendor capabilities and experiences in order to establish contracts and governance roles within the relationship. *Relationship* factors for the *client* organisation is about establishing and maintaining a shared information base, while the *vendor* organisation focuses on founding a detailed engagement plan with the client organisation managing client expectations.

4.3 Client-Vendor Knowledge Transfer Mechanisms for IS Outsourcing Govern Phase

The *govern* phase of IS outsourcing deals with the management of the IS outsourcing arrangement until it is terminated or renewed. Table 4 summarises the success of client-vendor knowledge transfer methods dependent on knowledge, client, vendor and relationship factors.

The main aim of *knowledge* factors for the *client* organisation is to maintain an explicit organisational knowledge repository, including the management of knowledge gaps that may arise. *Knowledge* factors for the *vendor* organisation focus knowledge transfer to the client organisation, closing knowledge gaps, and maintaining access to organisational knowledge bases and experts. *Client* factors for the *client* organisation point to ongoing maintenance of an up-to-date organisational knowledge base, creating and maintaining organisational skills to manage an outsource partner arrangement, detailed requirement specification and SLA management. *Client* factors for the *vendor* organisation include the communication of clear expectations and timelines for service delivery by business, the application of a consistent vendor management methodology across all aspects of the outsourcing agreement and the comparison of benefits and

Table 4. Vendor-client knowledge transfer methods for the IS outsourcing govern phase

Govern	Client organization	Vendor organisation
Knowledge factors	Mitigate any problems with regard to assumptions and addresses gaps of information and understanding Manage knowledge management repository, performance management standards and processes	Determine key knowledge gaps in the organisation and formalise how the knowledge in these areas would be transferred Transfer knowledge to client organisation in order to maintain service level agreement to the benefit of the organisation Ensure that internal resources now are able to complete a similar initiative independently Ensure key access to a formal repository with formal training and coaching sessions to all role players
Client factors	Train client organisation on outsource management requirements and skills Create common understanding of status between client organisation and the outsource vendor that could represented graphically through status dashboards, checklists, etc Monitor and assess benefit realisation within the project Conduct regular review of work and feed back Defined project governance process and procedure must include oversight and control of knowledge management flows Conduct outcome based analysis on usability of the end product produced, using key performance indicators set during the request for proposal (tender) phase Define detail design and requirements, measurable against standards, in order to measure effectively Establish client organisation monitoring capabilities as well establishing governance structures to manage the suppliers Conduct operational integration and expectation sessions discussing expectations and alignment between coming stages.	Define an initial stabilization period for service level agreement threshold definition Communicate clear expectations and timelines for service delivery by business Apply consistent vendor management methodology across all aspects of the agreement Compare benefits and outcomes to initial expected outcomes of the outsource and based on international best practice

(*Continued*)

Table 4. (*Continued*)

Govern	Client organization	Vendor organisation
	Implement well defined measureable business KPIs linked to well defined and measureable business processes. Build key client organisation skills to manage an outsource partner relationship, formal SLA's, strong relationship management/building and firm commercial knowledge	
Vendor factors		Specify clearly defined service level agreement at all hierarchical levels, containing measurements, monitoring, benefits, penalties and timeframes
Relationship factors	Foster wider collaboration and transparency, with detailed documentation	Provide regular feedback and create a client and vendor decision making body Set up a proper war room, with key stakeholders, team and vendor Create shared office space between internal project resources and outsourced project resources, allocated hand over time, joint documentation responsibilities Create a client organisation team that engages with the vendor as the vendor should not deal with business as usual/operational issues Pre-determined templates, containing agreed pertinent information which covers risk, monitoring and final delivery outputs, as well as defined commercial and management gates/phases with no room for ambiguity

outcomes to the initial expected outcomes of the outsource arrangement based on best practice. The specification of clearly defined SLA at all hierarchical levels, containing measurements, monitoring, benefits, penalties and timeframes, form part of *vendor* factors for the *vendor* organisation. *Relationship* factors for the *client* organisation refer to fostering wider collaboration and transparency underpinned by detailed documentation, while the *vendor* organisation focuses on regular feedback cycles to the client organisation, management of the arrangement including risks and issues, while steering clear of operational requirements and accountabilities.

5 Conclusion

Organisations have to manage knowledge in IS outsourcing activities, with emphasis on client-vendor knowledge transfer ensuring that no knowledge gaps are created through an IS outsourcing arrangement. In order to address the identified lack of applied research to explain how knowledge-vendor knowledge is transferred specifically in the context of IS outsourcing, this paper reported on a case study to determine client-vendor knowledge transfer methods in IS. The client-vendor mechanisms obtained were presented for each phase of IS outsourcing namely architect, engage and govern. Strong emphasis was placed on explicitly defining operational, system and business related knowledge bases. Specific actions were shared in order to identify people with key knowledge and implement specific plans to transfer their knowledge. The client-vendor knowledge transfer mechanisms obtained were considered in view of factors impacting the success of knowledge transfer in this context, namely, knowledge factors, client factors, vendor factors, and relationship factors.

From an organisational perspective, individual and organisational knowledge transfer must be considered in view of the capability that an outsource partner bring. In addition, the joint knowledge created by an organisation-vendor team, must be managed for competitive advantage and sustainable growth.

Since the interpretive case study was limited to a single major telecommunication organisation in South Africa and its associated outsource partners from multiple countries, further research is required in order to generalise the findings across the entire information telecommunication sector.

References

1. Dibbern, J., Goles, T., Hirschheim, R., Jayatilaka, B.: Information systems outsourcing: a survey and analysis of the literature. Data Base Adv. Inf. Syst. **35**(4), 6–102 (2004)
2. Teo, T.: Knowledge management in client-vendor partnerships. Int. J. Inf. Manage. **32** (2012), 451–458 (2012)
3. Hirschheim, R., Dibbern, J.: Information technology outsourcing: towards sustainable business value. In: Hirschheim, R., Heinzl, A., Dibbern, J. (eds.) Information Systems Outsourcing, pp. 1–19. Springer, Heidelberg (2014)
4. Rao, H., Nam, K., Chaudhury, A.: Information systems outsourcing. Commun. ACM **39**(7), 27–28 (1996)
5. Vilvovsky, S.: Differences between public and private IT outsourcing: common themes in the literature. In: 9th Annual International Digital Government Research Conference Montreal, Canada (2008)
6. Alexandrova, M.: IT outsourcing partnerships: empirical research on key success factors in Bulgarian organisations. Management **17**(2), 31–50 (2012)
7. Smuts, H., Kotze, P., Merwe, A.V.D., Loock, M.: Critical success factors for information systems outsourcing management: a software development lifecycle view. In: Proceedings of the Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists Bela Bela, South Africa (2010b)

8. Smuts, H., Kotzé, P., Merwe, A.V.D., Loock, M.: Information systems outsourcing issues in the communication technology sector. In: Proceedings of the IADIS International Conference Information Systems 2010. IADIS, Portugal (2010a)
9. Ranganathan, C., Balaji, S.: Critical capabilities for offshore outsourcing of information systems. MIS Q. Executive **6**(3), 147–164 (2007)
10. Narayanan, S., Swaminathan, J.M.: Information technology offshoring to India: pitfalls, opportunities and trends (2007). http://public.kenan-flagler.unc.edu/Faculty/swaminaj/research/paper/NS2.pdf. Accessed 14 Nov 2009
11. Fabriek, M., Brand, M.V.D., Brinkkemper, S., Harmsen, F., Helms, R.: Reasons for success and failure in offshore software development projects. In: 16th European Conference on Information Systems Galway, Ireland (2008)
12. Blumenberg, S., Wagner, H.-T., Beimborn, D.: Knowledge transfer processes in IT outsourcing relationships and their impact on shared knowledge and outsourcing performance. Int. J. Inf. Manage. **29**, 342–352 (2009)
13. Aydin, M.N., Bakker, M.E.: Analyzing IT maintenance outsourcing decision from a knowledge management perspective. Inf. Syst. Front. **10**, 293–305 (2008)
14. Bandyopadhyay, S., Pathak, P.: Knowledge sharing and cooperation in outsourcing projects: a game theory analysis. Decis. Support Syst. **43**, 349–358 (2007)
15. Beyah, G., Gallivan, M.: Knowledge management as a framework for understanding public sector outsourcing. In: 34th International Conference on System Sciences Hawaii. IEEE (2001)
16. Currie, W., Pouloudi, A.: Evaluating the relationship between IT outsourcing and knowledge management. J. Change Manage. **1**(2), 149–163 (2000)
17. Power, M.J., Desouza, K.C., Bonifazi, C.: The Outsourcing Handbook: How to Implement a Successful Outsourcing Process. Kogan Page Limited, London (2006)
18. Power, M.J., Bonifazi, C., Desouza, K.C.: The ten outsourcing traps to avoid. J. Bus. Strat. **25**(2), 37–42 (2004)
19. Sparrow, E.: Successful IT Outsourcing: From Choosing a Provider to Managing the Project. Practitioner Series. Springer, London (2003)
20. Tiwana, A., Bharadwaj, A., Sambamurthy, V.: The antecedents of information systems development capability in firms: a knowledge integration perspective. In: 24th International Conference on Information Systems (2003)
21. Hoffman, R., McDougall, P.: 2012 State of IT Outsourcing. InformationWeek, San Francisco (2012)
22. Brown, D., Fersht, P.: Executive Report: The State of Services and Outsourcing in 2014. KPMG and HfS Research (2014)
23. Borgia, E.: The internet of things vision: key features, applications and open issues. Comput. Commun. **54**(2012), 1–31 (2014)
24. Cullen, S., Willcocks, L.: Intelligent IT Outsourcing: Eight Building Blocks to Success. Computer Weekly Professional Series. D. Remenyi Butterworth-Heinemann, Oxford (2003)
25. Gellings, C.: Outsourcing relationships: the contract as IT governance tool. In: Proceedings of the 40th Hawaii International Conference on System Sciences Hawaii. IEEE (2007)
26. Aalders, R.: The IT Outsourcing Guide. Wiley, West Sussex (2001)
27. Dalkir, K.: Knowledge Management in Theory and Practice. Routledge, London (2013)
28. Gottschalk, P., Solli-Saether, H.: Knowledge transfer in IT outsourcing relationships: three international case studies. Int. J. Innov. Learn. **4**(2), 103–111 (2007)
29. Lee, M.K.O.: IT outsourcing contracts: practical issues for management. Ind. Manage. Data Syst. **96**(1), 15–20 (1996)

30. Balaji, S., Ahuja, M.K.: Critical team-level success factors of offshore outsourced projects: a knowledge integration perspective. In: Proceedings of the 38th Hawaii International Conference on System Sciences Hawaii. IEEE (2005)
31. Nickols, F.: The knowledge in knowledge management. In: Paper Commissioned for Knowledge Management Yearbook 2000–2001 (2001)
32. Clarke, T., Rollo, C.: Corporate initiatives in knowledge management. Educ. Training **43** (4/5), 206–214 (2001)
33. Nonaka, I., Toyama, R., Byosiere, P.: A theory of organisational knowledge creation: understanding the dynamic process of creating knowledge. In: Dierkes, M., Antal, A.B., Child, J., Nonaka, I. (eds.) Handbook of Organizational Learning and Knowledge, pp. 491–517. Oxford University Press, New York (2001)
34. Nonaka, I., Takeuchi, H.: The Knowledge Creating Company. Oxford University Press, Oxford (1995)
35. Nguyen, P.T., Babar, M.A., Verner, J.M.: Critical factors in establishing and maintain trust in software outsourcing relationships. In: ICSE. ACM (2006)
36. Sabherwal, R.: The role of trust in outsourced IS development projects. Commun. ACM **42** (2), 80–87 (1999)
37. Benamati, J., Rajkumar, T.M.: A design of an empirical study of the applicability of the technology acceptance model to outsourcing decisions. In: SIGCPR. ACM (2002)
38. Al-Salti, Z.: Knowledge Transfer and Acquisition in IS Outsourcing: Towards a Conceptual Framework, p. 11. Brunel Business School, Brunel University, London (2009)
39. Yam, R., Tang, E.: Knowledge assets, supplier and customer coordination and product performance. In: Management, Knowledge and Learning International Conference 2013, Knowledge Management & Innovation, Zadar, Croatia (2013)
40. Srinivas, H.: Position paper on knowledge asset management, Sept 1999. http://www.gdrc.org/kmgmt/km-1.html
41. Mansingh, G., Osei-Bryson, K.-M., Reichgelt, H.: Building ontology-based knowledge maps to assist knowledge process outsourcing decisions. Knowl. Manage. Res. Pract. **7**, 37–51 (2009)
42. Choy, C.S., Yew, W.K., Lin, B.: Criteria for measuring KM performance outcomes in organisations. Ind. Manage. Data Syst. **106**(7), 917–936 (2006)
43. Olivier, M.S.: Information Technology Research: A practical Guide for Computer Science and Informatics, 2nd edn. Van Schaik Publishers, Pretoria (1997) (Edition 2004)
44. Larsen, M.A., Myers, M.D.: When success turns into failure: a package driven business process re-engineering project in the financial services industry. J. Strateg. Inf. Syst. **8**, 397–417 (1999)
45. Yin, R.K.: Case Study Research: Design and Methods, 3rd edn. Sage Publications Inc., Thousand Oaks (2003)
46. Welman, C., Kruger, F., Mitchell, B.: Research Methodology, 3rd edn. Oxford University Press, Cape Town (1994)
47. Flick, U.: Designing Qualitative Research. The SAGE Qualitative Research Kit, Sage Publications Ltd, Thousand Oaks (2007)
48. Leedy, P.D., Ormrod, J.E.: Practical Research: Planning and Design, 10th edn. Pearson Education Limited, New Jersey (2014)

An Empirical Study of the Effect of SCM Practice on Corporate Performance (Based Specifically on the Chinese Manufacturing Industry)

Baolin Xia(✉) and Akira Kamoshida

Graduate School of International Management, Yokohama City University, 22-2 Seto, Kanazawa-ku, Yokohama 236-0027, Japan
xiabaolinlin@gmail.com, mail@akirakamoshida.jp

Abstract. Few empirical studies on the practical effectiveness of supply chain management (SCM) in Chinese corporations have been conducted. Given this situation, using the Chinese industry as the research target, the current status of SCM practice and its practical effectiveness in China are investigated herein based on the data analysis of well-designed questionnaires. The contribution level of SCM practice to corporate business performance is examined through the construction and estimation of structure equation models. The results of this empirical study could be used as a guidance for the SCM practice in the Chinese manufacturing industry in the future.

Keywords: Supply chain management (SCM) · Strategic supplier partnership (SSP) · Customer relationship (CR) · Information sharing (IS) · Lean production system (LPS) · Corporate performance (CP) · Structure equation models (SEM)

1 Introduction

In recent decades, economic globalization has boomed, leading to fierce competition among enterprises, deregulation, and the rapid advancement of information technology. Under these circumstances, companies strive to achieve optimal integrated management in the supply chain, from source suppliers to terminal consumers (with the exception of the inner optimization of some individual companies). Supply Chain Management (SCM) is often proposed as a solution. In particular, the successful application of SCM in companies such as Dell Computer, Inc., Wal-Mart, Inc., and Hewlett-Packard Company speaks for itself. However, the effect of SCM on Chinese companies is little understood thus far.

Due to the traditional planned economy that prevailed in past decades in China, which has been hailed the "world's factory," Chinese distribution systems were restricted and, therefore, remained underdeveloped. Recently, the situation has changed drastically. Owing to the transition from a seller's market to a buyer's market due to the competitive situation, the optimization of logistics management – namely, reducing the cost of logistics, improving CS, and optimizing efficiency in terms of transportation lead time for stock – becomes a great challenge that needs urgently to be conquered by

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 120–133, 2015.
DOI: 10.1007/978-3-319-21009-4_10

Chinese companies. SCM is thus becoming increasingly important for Chinese corporations in such competitive market circumstances.

However, many people and organizations remain doubtful about the validity of SCM. Such skepticism is based mainly on the experiences of particular companies that have found reasons to question the effectiveness of SCM. This uncertainty is related to the question of whether or not SCM is as valid in the Chinese context as it is for European companies. As previous empirical studies on SCM are lacking, it is necessary to conduct further research, therefore, into the effectiveness of SCM.

Based on the above-mentioned background, the purpose of this study is to assess the conditions for the application of SCM in the Chinese manufacturing industry and thus to clarify the relationship between SCM and business performance.

2 Literature Review

A comprehensive literature review of the topic indicates that the effective use of SCM is an important method for improving corporate performance (CP). Some researchers have split SCM into several components before studying the effect of these components on CP. Frohlich and Westbrook (2001), for example, have analyzed how the integration of suppliers and customers affects CP [1]. Eriksson (2010) brings the idea of "lean thinking" into his study of the relationship between SCM and CP, using a case-study method for his research [2]. Tan et al. (1998) focus on the effects of suppliers and customers, which are the two key components of SCM, on CP [3]. Kim et al. (2014) show that the coordination and cooperation of supply chain integration and diversification strategies have a significant effect on both the whole SCM performance and the firm's performance through simultaneous equation analysis on the proposed model representing structural relationships among four constructs by employing LISREL [4].

Other researchers divide the concept of CP into several parts, examining the impact of SCM practice upon them separately. Chan et al. (1997) point out that SCM could be used with information systems to influence many aspects of CP, such as the performance of the information systems themselves, which might lead to a competitive advantage for companies [5]. McLaren et al. (2004) have used path analysis to look at sales volumes in order to specify and measure CP. They found that effective SCM practice performed by a corporation can have a strong effect on corporate market performance [6]. The use of a Lean System involves the practice of driving out all unnecessary costs, time, and other waste from the entire supply chain. Here, Mason and Towill (1997) have found that lean thinking and practice has become a very important dimension for the successful implementation of SCM [7]. Arlbjorn et al. (2011) outline a model that illustrates under which conditions lean is deemed most appropriate according to the type of service delivered [8].

Compared with the research conducted abroad, quantitative research into the relationship between SCM and CP began late in China; in the main, it has been carried out only in the last few years. Wang and Lyu (2010) have studied the relationship between financial SCM and CP in terms of the financial supply chain. They divided their samples between the physical and the financial supply chain by adopting the quantitative method, before assessing the relationship between SCM and CP via a linear regression. The results showed that physical SCM has no obvious effect on CP, whereas financial SCM has a direct influence [9].

Wang and Zhang (2007) have studied financial indices, such as “return on investment,” “return on equity,” and “return on assets,” for listed corporations. In all the listed corporations targeted by this study, an Enterprise Resource Planning (ERP) system has been implemented. The researchers’ results indicated that the ERP system has a clear effect on CP, especially in small corporations. They also found that the influence on CP is more apparent in non-production corporations [10]. Zhang and Huang (2012) show that corporate governance has a positive moderating effect on the relationship between ERP investments and firm performance, and that firms with a higher level of diversification perform worse utter ERP implementation based on the empirical study on the samples of 137 listed companies in the US from 1998 to 2007 [11].

Feng (2012) discusses the relationship between SCM strategy, information system strategy, and CP in his Master’s thesis. He studied the relationships between these three factors quantitatively by issuing questionnaires and analyzing the returned sample datasets using structural equation model estimation. The results showed that, in environments where full information systems are used, CP can be improved significantly through the adoption of a hybrid supply chain [12].

Ye (2009) has conducted an in-depth study into the relationship between four factors: trust between supply chain partners, relationship commitment, information sharing, and operation performance. The results indicated that trust between supply chain partners has positive effect on the other three factors, whereas relationship commitment had almost no effect. Further, the quality and level of information sharing was found to have an effect on operation performance [13]. Ding et al. (2014) applied stepwise multiple regression analysis in their study to test the hypothesized relationships based on a survey questionnaire to 600 Australian beef processors. And the findings show that strategic alliance, information quality and trust and commitment are significantly related to food quality. In particular, the standardized coefficient shows that information quality has a significant positive relationship with food quality [14].

Chen (2007) focuses on the effects of supply chain integration levels on CP within production corporations, which is an important component of SCM. The study was conducted using quantitative analysis methods, such as exploratory factor analysis and linear regression. The research results showed that there is a positive relationship between supply chain integration and CP [15]. Supply chain integration is a multi-dimensional construct, including the four dimensions strength, scope, duration, and depth of integration. Eriksson (2015) conducted a multiple case study of four engineering projects in their study and the empirical findings indicate that these four dimensions are critical when conceptualizing and implementing partnering in engineering projects. The results show that there are strong interdependencies among the four dimensions, suggesting that it is crucial to manage them simultaneously and systemically rather than in isolation [16].

3 Proposed Hypothesis SEM Models

Structural Equation Modelling (SEM), or path analysis, is a very powerful multivariate technique that enables researchers to measure direct and indirect effects. SEM is often used to perform test models with multiple dependent variables and also when utilizing several regression equations simultaneously.

The main purpose of this study is to reveal the effects of SCM practice on CP in the Chinese manufacturing industry. Based on this purpose and on the arguments listed in the above literature review, two hypotheses are proposed. First, Hypothesis H1 is proposed and will be tested using Hypothesis SEM Model 1.

Hypothesis H1: Supply Chain Management (SCM) practice has a significant effect on CP.

It is easier to estimate the role of SCM practice by examining 4 dimensions separately, which allows the degree of their contributions to CP to be analyzed. Therefore, in order to decompose SCM practice into four dimensions, the following four sub-hypotheses (Hypotheses 2–5) are proposed, which will be tested using Hypothesis SEM Model 2.

Hypothesis H2: The customer relationship (CR) has a significant effect on CP.
Hypothesis H3: Information Sharing (IS) has a significant effect on CP.
Hypothesis H4: Strategic Supplier Partnerships (SSP) have a significant effect on CP.
Hypothesis H5: Lean Production Systems (LPS) have a significant effect on CP.

The hypotheses listed above are elaborated further in Figs. 1 and 2, in which the following notation is used:

- ξ represents latent exogenous variables and X represents their observable indicators.
- Y represents the indicators of the latent endogenous variables, which are symbolized by η.
- The measurement errors associated with the endogenous indicators are represented by ε, while δ represents measurement errors in the exogenous indicators, and ζ is used for structural disturbances.

The structural and measurement sub-models are formulated as follows:

$$\begin{aligned} \boldsymbol{\eta} &= \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \\ \mathbf{Y} &= \boldsymbol{\Lambda}_{\mathbf{Y}}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \\ \mathbf{X} &= \boldsymbol{\Lambda}_{\mathbf{X}}\boldsymbol{\xi} + \boldsymbol{\delta}, \end{aligned}$$

where $\boldsymbol{\eta}$ is the latent endogenous variables vector, $\boldsymbol{\xi}$ is the latent exogenous variables vector, $\mathbf{Y}$ is the indicators vector for latent endogenous variables, $\mathbf{X}$ is the indicators vector for latent exogenous variables, $\boldsymbol{\varepsilon}$ represents the measurement errors in endogenous indicators, $\boldsymbol{\delta}$ represents the measurement errors in exogenous indicators, $\boldsymbol{\zeta}$ is the structural disturbances vector, $\mathbf{B}$ is the structural parameters matrix for latent

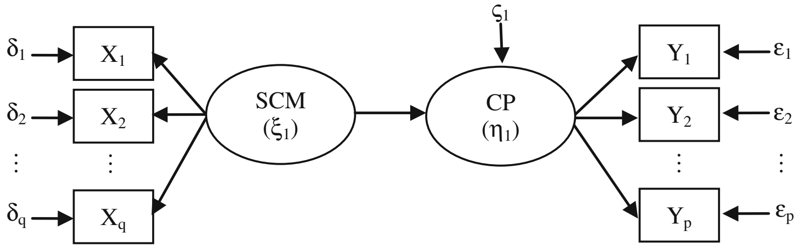


Fig. 1. Hypothesis SEM model 1

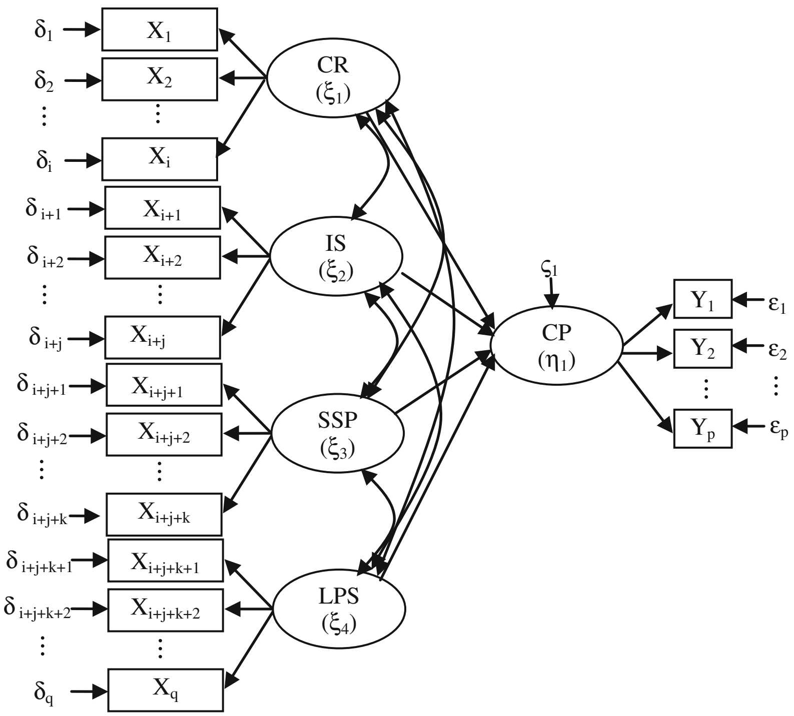


Fig. 2. Hypothesis SEM model 2

endogenous variables, $\mathbf{\Gamma}$ is the structural parameters matrix that relates latent endogenous to exogenous variables, and $\mathbf{\Lambda_X}$ and $\mathbf{\Lambda_Y}$ are factor loadings matrices relating indicators to latent variables.

In order to compile the model, some of the parameters in $\mathbf{B}$, $\mathbf{\Gamma}$, $\mathbf{\Lambda_X}$, and $\mathbf{\Lambda_Y}$ must be constrained, which is achieved typically by setting the parameters to 1 (as is the case here). In this study, the process of Structural Equation Modelling was performed using the statistical software SPSS Amos. The estimation details will be described in the next section.

4 Empirical Study

4.1 Study Approach and Variables Selection

In this study, the hypothesis model is verified using structural equation modelling based on the data collected from questionnaire surveys. The questionnaire survey is an important method for evaluating current SCM level within corporations. In our study, to avoid the names and private information of corporations being leaked externally, confidentiality agreements were signed before any data was collected. Additionally, the corporations participating in the questionnaire surveys were offered feedback on the results as an incentive. In February 2014, the questionnaire surveys were given to the China Logistics Association [17] and were then issued to the corporations belonging to the association so that a mass of data could be collected for the study.

In terms of questionnaire design, the questions were modeled in part on those that have been used in survey questionnaires by Li (2002), which comprise three main components [18]:

(1) **Face Sheet:** this sheet gives basic information about the corporation participating in the survey, such as the corporation name, business division, industry type, volume of business, and number of employees
(2) **SCM Practice:** this element involves 4 main items: (1) The customer relationship, (2) Information Sharing, Strategic Supplier Partnership, Lean Production System. For our study, these items were divided into sets of 5, 4, 3, and 3 sub-items respectively, thus totally 15 evaluation items.
(3) **CP:** this element includes four evaluation items: (1) investment income, (2) increment rate of investment income, (3) Market shares, (4) increment rate of market shares. These items were adopted as the CP variables in our study to allow measurement of the effectiveness of SCM practice. The participating corporations were asked to rate their CP in relation to other corporations in the same type of industry in terms of the 4 evaluation items above. A 5 point scale was used: (1) Strongly Disagree, (2) Disagree, (3) Neutral, (4) Agree, (5) Strongly Agree. The questionnaires were issued to the participant corporations in two ways: either by a web version posted on a specific web page or a paper version sent in the mail.

In addition, this study restricted the survey targets to production corporations in China because high-level SCM practices can usually be found in the Chinese manufacturing industry.

4.2 Reliability, Validity Test and Factor Analysis

For the pilot study, 160 questionnaire sheets were issued, with 92 sheets being returned ultimately. Reliability and validity tests were performed for the 92 questionnaires using the SPSS statistical package (version 21). Based on these evaluation criteria, the questionnaire sheets were modified and final versions were created for the main, formal survey, which was conducted in across 600 corporations. 250 questionnaire sheets were returned (the response rate was about 42 %), of which 240 sheets were valid. The results of the analysis using SPSS 21 are shown in Tables 1 and 2.

Table 1. Factor loading matrix after varimax rotation

Evaluation Items	Factors			
	CR	IS	SSP	LPS
1-1) We frequently measure and evaluate customer satisfaction	.863	.229	.137	-.087
1-2) We periodically evaluate the importance of our relationship with our customers	.817	.065	.240	-.021
1-3) We frequently determine future customer expectations	.767	-.011	.114	.124
1-4) We frequently evaluate the formal and informal complaints of our customers	.739	.291	.231	-.125
1-5) We frequently interact with customers to set reliability, responsiveness, and other standards	.732	.246	.124	.181
2-1)We share our business units' proprietary information with trading partners	.131	.858	.176	.047
2-2) We inform trading partners in advance of changing needs	.181	.828	-.064	.144
2-3) Our trading partners share proprietary information with us	.147	.812	.176	.142
2-4) Our trading partners keep us fully informed about issues that affect our business	.121	.749	-.081	.098
3-1) We consider quality as our number one criterion in selecting suppliers	.247	-.100	.822	-.026
3-2) We regularly solve problems jointly with our suppliers	.406	.045	.805	-.010
3-3) We include our key suppliers in our planning and goal-setting activities	.078	.176	.588	.141
4-1) Our firm uses a "Pull" production system	.028	.118	.122	.822
4-2) We involve our customers in process/product design	.075	.113	-.163	.737
4-3) Our firm reduces set-up times	-.042	.095	.149	.705
Cronbach's Alpha	0.875	0.850	0.703	0.666
Eigen Value	3.408	2.922	1.964	1.861
Cumulative Contribution Ratio (%)	22.719	42.198	55.291	67.694

Table 2. Factor analysis for CP

Evaluation items	Corporate performance	Cronbach's alpha	Eigen value	Cumulative contribution ratio (%)
1-1) Investment income	.854	0.856	2.843	71.084
1-2) Increment rate of investment income	.932			
1-3) Market share	.827			
1-4) Increment rate of market share	.749			

First, the reliability of the questionnaire was tested by assessing its internal consistency, which was measured using Cronbach's alpha measure. Internal consistency refers to the degree of correlation between the items on a scale. It is expected that the items in each of the domains of the questionnaire should correlate moderately with each other, but should also contribute independently to the overall score for the domain. A perfect correlation of 1.0 suggests that questions are being used to measure almost identical constructs, which results in item redundancy, whereas a poor correlation suggests that the items may be testing for a number of different traits. It has been suggested by Nunnally (1978) that an alpha value of ≥0.7 is acceptable [19], although Carmines and Zeller (1979) [20] recommend values >0.8. As the test results given in Table 1 show, the internal consistency for our questionnaire (tested using Cronbach's alpha) exceeds the minimum requirement of 0.7 for reliability across all the domains, except LPS, which is slightly lower than 0.7. These results demonstrate that the internal consistency of the questionnaire is acceptable.

Next, factor analysis was conducted in order to test the validity of the questionnaire. Before the factor analysis, Kaiser-Meyer-Olkin's (KMO) Test of Sample Validity and Bartlett's Test of Sphericity were performed. The KMO value of a questionnaire should be ≥0.7, with a very minimum of ≥0.5. In statistics, Bartlett's Test of Sphericity checks whether the observed correlation matrix for the evaluation items diverges significantly from the identity matrix (theoretical matrix under H0: the variables are orthogonal). Factor analysis can compress the available information only if we reject the null hypothesis ($p < 5$ %). According to the results of the analysis for the sample dataset, the KMO index was 0.714. We rejected the null hypothesis at the 5 % level (p-value = 0.000). It was determined, therefore, that we could perform a factor analysis effectively using our dataset.

The factor analysis results show that the first four factors represent 67.694 % of the available variance across 15 items. These four factors indicate that we have an accurate picture of the information available within our dataset. A varimax rotation was used to maximize the sum of the variances of the squared loadings (squared correlations between variables and factors). The calculation on the sample dataset converged after 5 iterations. Most of the factor loadings after rotation were >0.7, which means the convergent and divergent validity of the adopted questionnaire is acceptable.

4.3 Model Fit Test for SEM

The model fit indices determine how well a priori SEM model fits the sample data and demonstrate which proposed model is most suitable [21]. These measures provide a fundamental indication of how well the proposed theory suits the data. The calculations do not rely on comparisons with a baseline model; instead, they measure how well the model fits in comparison with having no model at all [22]. The 4 frequently used indices and the corresponding criteria are listed in Table 3.

Table 3. Model fit indices and criteria [23]

Index	Possible range	Acceptable range	Unacceptable range
Chi-square	≥0	Judged by p value	Judged by p value
χ^2/df	<2	1 ~ 2	–
RMSEA	RMSEA ≥ 0	<0.05	>0.1
GFI	GFI ≤ 1	>0.90	<0.9
CFI	0 ≤ CFI ≤ 1	>0.90	<0.9

In this study, Structural Equation Modelling was performed using the statistical software SPSS Amos (version 21) and the estimation results are shown in Figs. 3 and 4. An annotated fit summary is displayed in Table 4. For Hypothesized Model 1, the χ^2/df value is 1.928 (between 1 and 2). This indicates that Hypothesized Model 1 cannot be rejected. The CFI index is close to 1, which indicates that the model is an excellent fit. Similarly, the GFI index is also nearly 1, again suggesting an excellent fit. The RMSEA value is 0.032, which is smaller than the conventional 0.05 value for an acceptable fit. Similarly, for Hypothesized Model 2, according to the criteria listed in Table 3, all the model fit indices are quite satisfactory, which shows that Hypothesized Model 2 also represents the sample data very well and can be used, therefore, for testing the hypotheses in the following step.

Table 4. Results for model fit indices

	χ^2/df	CFI	GFI	RMSEA
Model 1	1.928	0.968	0.925	0.032
Model 2	1.725	0.918	0.885	0.041

4.4 Hypothesis Tests

The path diagram for Hypothesis Model 1, as estimated using SEM, is shown in Fig. 3. The same is shown for Hypothesis Model 2 in Fig. 4. The hypothesis test results are displayed in Table 5. For hypothesis H1, CP had a path coefficient of 0.778 from the factor of SCM practice, and the corresponding T value was 4.764. The T value of the estimations is larger than the conventional value of 1.96 for accepting a 0.05 significance level. This indicates that hypothesis H1 cannot be rejected at a 0.05 significance level. It is clear that the factor of SCM practice had a strong effect on the indicator of CP, for the path coefficient of 0.778 is quite large.

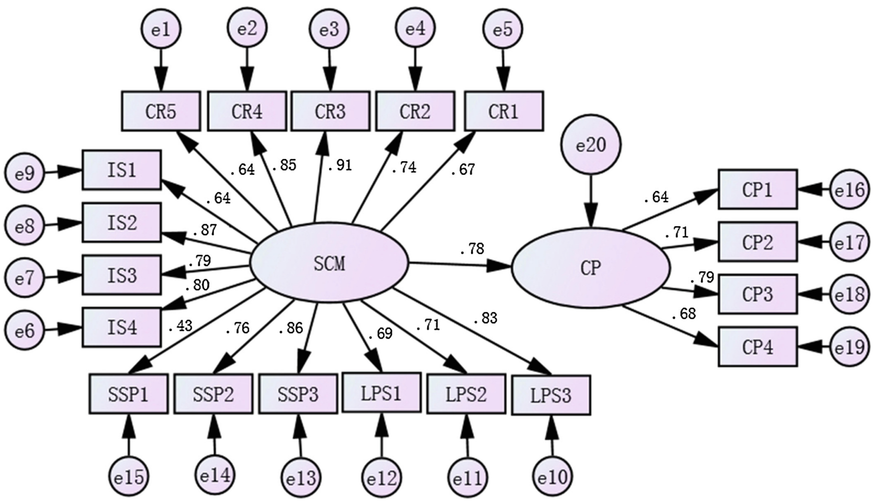


Fig. 3. Path diagram for Hypothesis model 1

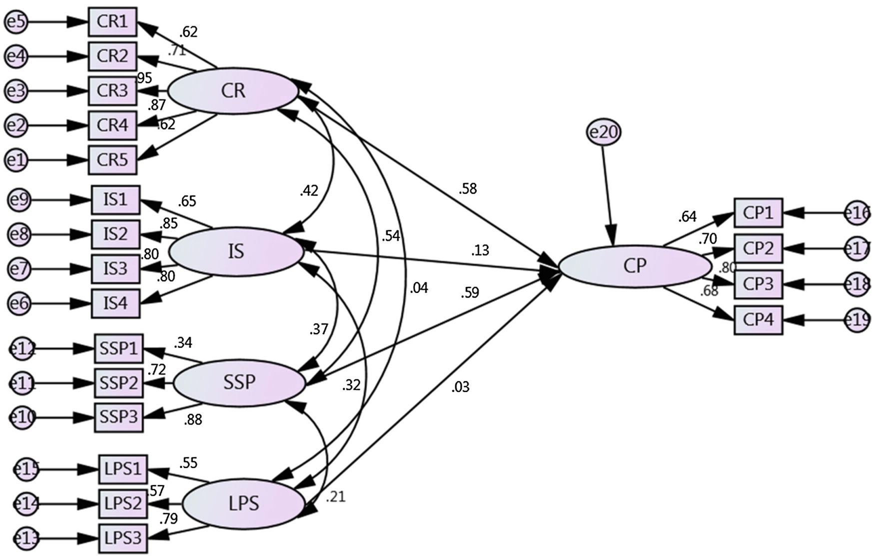


Fig. 4. Path diagram for Hypothesis model 2

Table 5. Hypothesis test results

Hypothesis	Standardized path coefficient	T value	Test result
H1 SCM→CP	0.778	4.764	Not rejected
H2 CR→CP	0.577	5.406	Not rejected
H3 IS→CP	0.134	0.698	Rejected
H4 SSP→CP	0.589	2.529	Not rejected
H5 LPS→CP	0.029	0.375	Rejected

Similarly, for hypothesis H2, CP had a path coefficient of 0.577 from the factor of the customer relationship (CR), and the corresponding T value was 5.406. For hypothesis H4, CP had a path coefficient of 0.589 from the factor of Strategic Supplier Partnership (SSP), and the corresponding T value was 2.529. These T values larger than 1.96 indicate that hypotheses H2 and H4 cannot be rejected at a 0.05 significance level. The path coefficients show that the factors of customer relationship (CR) and Strategic Supplier Partnership (SSP) have a strong impact on the indicator of CP. On the contrary, for hypotheses H3 and H5, CP had a path coefficient of 0.134 from the factor of Information Sharing (IS) and 0.029 from the factor of the Lean Production System (LPS), with corresponding T values of 0.698 and 0.375 respectively. The low T values indicate that hypotheses H3 and H5 can be rejected. This means that the factors of Information Sharing (IS) and the Lean Production System (LPS) have no clear effect on CP.

5 Conclusions

The results of the hypothesis test show that there is a correlation between SCM practice and CP, as shown in Fig. 3.

The estimation results for hypothesis test H1 show that it cannot be rejected. This indicates that SCM has a positive influence on CP in the manufacturing industry. It also demonstrates that the 4 measurement dimensions selected for the study (Strategic Supplier Partnership, the customer relationship, Information Sharing, and the Lean Production System) are appropriate measurement dimensions for SCM practice. Indeed, the results suggest that the SCM practice formed by these four dimensions is able to have a positive effect on CP.

Firstly, a SSP aims to build a long-term, cooperative relationship, involving coordination and collaborative problem solving with core suppliers. Suppliers are an important aspect of the competitive advantages that a company can gain. There are many advantages in building strategic partnerships with suppliers in the manufacturing industry. For instance, the launch time for products can be shortened significantly, the quality of the products can be enhanced, and the market share and customer satisfaction can increase.

Secondly, customer relationship management is intended to improve customer satisfaction, which allows harmonious relationships to be created and maintained with customers. If a company can keep close and long-term relationships with customers, their market risks will be reduced greatly, and customer satisfaction and loyalty will grow. Moreover, the profits of the company will therefore increase.

Thirdly, Information Sharing (IS) involves the mutual sharing and transfer of information between SC members. It is possible to suppress the bullwhip effect by integrating management systems with distribution; for example, required information should be shared between teams and an inventory management system should be created. Furthermore, to improve the performance of the SC overall, it is necessary to build a mechanism for solving problems internally by sharing information mutually.

Finally, the Lean Production System (LPS) aims to eliminate the waste in the manufacturing process. In particular, Japanese automotive companies use lean manufacturing and development to reduce the total costs of production systematically in an attempt to establish the nation as a world-class player in the industry. The advantages of this system lie in productivity, manufacturing quality, development productivity and speed, product competitiveness, cost, and quality, which contribute significantly to the competitive advantage of a company and its profits.

Thus, in the process of SCM practice, a focus should be placed on the effective use of these 4 factors, which can help to improve the market performance and financial performance of a corporation. This study has also shown that the research methods for studies of SCM practice in Europe and the USA are also suitable for use in relation to Chinese corporations.

The estimation results show that hypothesis tests H2 and H4 cannot be rejected, either, as shown in Fig. 4. This indicates that the customer relationship and a Strategic Supplier Partnership have a positive effect on CP in SCM practice.

Contrastingly, the estimation results in Fig. 4 indicate that hypothesis test H3 and H5 can be rejected. Numerous literatures have shown that IS and LPS are the key elements that are essential to the SCM practices. Normally the higher performance of IS and LPS should have a positive correlation with CP. However, the results of this empirical study show that there is no clear correlation between IS, LPS and CP in Chinese companies. It means the implementations of IS and LPS are invalid for improving the CP of Chinese companies in the SCM practice.

Now we will firstly try to give some explanations to the invalidity of IS in Chinese companies. For the sake of reducing the total lead time, IS in SCM should optimize and improve the efficiency of the entire business process, rather than a single company, including all the shareholders outside the company, such as logistic suppliers and retailers, etc. It is essential to share information throughout the supply chain. Sharing information among enterprises in China has been recognized as an important factor of SCM, however, it is believed that information sharing among companies cannot get a clear effect on CP for the low level of the enterprise management. In addition, most of the Chinese companies started to pay attention to the company information in the 1990s, but many of them still don't have a clear intention to information sharing yet. Without a good understanding of the significance of information sharing between companies, most of the Chinese companies only followed the trend, and created a large number of invalid systems which cannot be used at all. Thus, in SCM practice information sharing cannot make much contribution to the corporate performance.

On the other hand, discussions with some of the corporations' managers revealed that cooperation between different corporations in a supply chain is mainly established on the basis of contract. Both parties involved keep closely to the contract, communicating and transferring information based on their respective needs only. In other words, all the information that is shared is selective and asymmetric. In addition, although a mass transfer of information may occur, little feedback is given or received, which leads to information distortion. In the process of SCM practice, therefore, corporations should intensify the extent of Information Sharing with supplier partners, transferring information in a timely manner and avoiding sharing vague, unnecessary information. In this way, the quality and effectiveness of Information Sharing between production corporations and supplier partners would be improved.

Finally, we consider the reason why LPS has no significant effect on the corporate performance in Chinese companies. Generally speaking, big problems may be existing in the following three aspects: concept revolution of top management in the organization, basic understanding of LPS technique, and the change of enterprise culture. The literature review shows that the success of LPS implementation requires the following three key points managed simultaneously and systemically.

The first key point is unyielding commitment to the top revolution. Just-in-time (JIT) and "Automation" are considered to be the two pillars of LPS. Strictly speaking, the automation start first followed by the development of JIT. JIT is the final goal, and automation is the means to achieve this goal. The implementation of JIT requires the concept revolution of top management in the organization.

The second key point is the basic understanding of LPS techniques. For the companies promoting to introduce LPS into their SCM system, it is really necessary for them to accurately understand the LPS techniques and their relationship between each other. In general, LPS techniques can be summarized as following: Kanban system, levelling (small-lot), automation, multi-functionalization, standardization (continuously improving work standards), improvement of layout, and the single setup. For stabilizing LPS, all the above mentioned techniques should be accurately understood and mastered. Particularly in many of the Chinese companies, those techniques have not been understood and mastered well so far.

The final key point is "continuous improvement" and "learning organization". As described above, to introduce and implement TPS, the first condition is the concept revolution related to the ultimate goal of top management, and the unyielding determination and declaration to run this revolution. And the second condition is the accurate understanding of the various techniques of LPS, as well as the relationships between them. However, it is insufficient for stabilizing the LPS to meet only these two conditions. The third condition is the change of enterprise culture. Specifically, it is necessary to construct a enterprise culture with a character of "continuous improvement" and "learning organization". After the three conditions mentioned above are all satisfied, it is necessary to begin to shorten the production period. For stabilizing the introduction and implementation, the production period should be shortened at a breakthrough level, yet no turning back must also be guaranteed as a mechanism.

When you look into the conditions described above and consider the present situation of China, without the three preconditions — unyielding commitment to the concept revolution of top management in the organization, basic understanding of LPS techniques, "continuous improvement" and "learning organization", the implementation of LPS does not go well in Chinese enterprises at the present situation.

Considering the present situation of Chinese enterprises specifically, the invalidity of the Lean Production System in Chinese enterprises may be due to two reasons: first, corporations may fail to understand and implement the Lean Production System systematically and practically; second, corporations may lack the ability to combine the Lean Production System organically with management mechanisms. Its implementation may also be affected by personnel capability. The ultimate purpose of production management is to realize the balanced development of five key goals: safety, quality, cost, delivery, and people. Traditional lean production theory offers a set of principled frames and common tools, which corporations should adopt and adapt to their own conditions, designing a Lean Production System with production management specifically in mind. Each corporation has its own strategic focuses, special production contents, and production condition limitations, and so lean production methods should be designed flexibly, according to a company's own strategy, environment, and personnel, instead of the general lean tools and methods being imitated indiscriminately or the practices of other corporations being utilized.

References

1. Frohlich, M.T., Westbrook, R.: Arcs of integration: an international study of supply chain strategies. J. Oper. Manage. **19**(2), 185–200 (2001)
2. Erik Eriksson, P.: Improving construction supply chain collaboration and performance: a lean construction pilot project. Supply Chain Manage. Int. J. **15**(5), 394–403 (2010)
3. Tan, K.C., Kannan, V.R., Handfield, R.B.: Supply chain management: supplier performance and firm performance. Int. J. Purchasing Mater. Manage. **34**(3), 2–9 (1998)

4. Kim, S.W., Kim, T., Choi, K.H., Yang, J.A.: Comparative analysis on the diversification effects between firm performance and whole supply chain performance in Korea and Japan. J. Korea Trade **18**(1), 1–27 (2014)
5. Chan, Y., Huff, S., Barclay, D., Copel, D.: Business strategic orientation, information systems strategic orientation, and strategic alignment. Inf. Syst. Res. **8**(2), 125–150 (1997)
6. McLaren, T.S., Head, M.M., Yuan, Y.: Supply chain management information systems capabilities: an exploratory study of electronics manufacturers. IseB **2**(2–3), 207–222 (2004)
7. Mason, J.K., Towill, D.R.: Information enrichment: designing the supply chain for competitive advantage. Supply Chain Manage. **2**(4), 137–148 (1997)
8. Arlbjorn, J.S., Freytag, P.V., de Haas, H.: Service supply chain management: a survey of lean application in the municipal sector. Int. J. Phys. Distrib. Logistics Manage. **41**(3), 277–295 (2011)
9. Wang, D.M., Lyu, B.F.: An empirical study on the impact of financial supply chain management on short-term corporate performance. Math. Pract. Theory **40**(2), 57–65 (2010)
10. Zhang, J.D., Wang, L.Y.: Connection between firm's performance and ERP implementation - based on data of China listed companies. Manage. World **2007**(3), 117–136 (2007)
11. Zhang, L., Huang, J.H.: The moderating factors in the relationship between ERP investments and firm performance. J. Comput. Inf. Syst. **53**(2), 75–84 (2012)
12. Feng, Y.Y.: The Study on the Relationship Among Information Systems Strategy, Supply Chain Management Strategy and Supply Chain Performance. Jilin University, Jilin (2012)
13. Ye, F., Li, Y.N., Zhang, H., Lei, X.Y.: Relationship among influential factors and level of information sharing and enterprise performance in supply chains. Chin. J. Manage. **016**(6), 743–750 (2009)
14. Ding, M.J., Jie, F., Parton, K.A., Matanda, M.J.: Relationships between quality of information sharing and supply chain food quality in the Australian beef processing industry. Int. J. Logistics Manage. **25**(1), 85–108 (2014)
15. Chen, X.H.: Impact of Supply Chain Integration on Manufacturing Enterprise's Performance. Hunan University, Hunan (2007)
16. Eriksson, P.E.: Partnering in engineering projects: four dimensions of supply chain integration. J. Purchasing Supply Manage. **21**(1), 38–50 (2015)
17. China Guangdong Logistics Association originally founded in 2006, is an association of the companies, business groups and relevant economic organizations involved in the logistics industry, which can be joined freely and is legally independent. Currently, the association is composed of more than 1,600 members and 197 staff
18. Li, S.H.: An Integrated Model for Supply Chain Management Practice Performance and Competitive Advantage. The University of Toledo, Toledo (2002)
19. Nunnally, J.C.: Psychometric Theory. McGraw Hill, New York (1978)
20. Carmines, E., Zeller, R.: Reliability and Validity Assessment: Quantitative Applications in the Social Science. Sage, Beverley Hills (1979)
21. McDonald, R.P., Ho, M.H.: Principles and practice in reporting statistical equation analyses. Psychol. Methods **7**(1), 64–82 (2002)
22. Jöreskog, K., Sörbom, D.: LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language. Scientific Software International Inc., Chicago (1993)
23. Asano, H., Suzuki, T., Kojima, T.: Introduction to Practical Structural Equation Modelling, p. 28. Kodansha Ltd., Tokyo (2005)

Impact of Knowledge Sharing on Decision Making and Situation Analysis

Taher ALBraik ALAmeri(✉)

Faculty of Engineering and IT, British University in Dubai, Dubai, United Arab Emirates
toti7445441@gmail.com

Abstract. The research discusses the role on knowledge management on improved decision making skills of the managers and the business. Knowledge is considered as power and the present century has also been coined as the century of knowledge, all great nations are striving to develop a knowledge economy, these proves that knowledge management be a vital for organizations too. Access to knowledge at the right time is critical, it can help the managers make the right decision promptly and swiftly, especially when there is a lot of unneeded data or information overload.

Keywords: Knowledge management · Organizational decision making

1 Introduction

'Decision Making' is an ultimate term for all business professionals, the world of business is the world of decisions. In the corporate world employees are paid to make decisions [9]. People are evaluated in the corporate world on the basis of their decision making skills, the rate of success of decision making is also connoted as the rate of success of the enterprise and the professional. The term decision making was devised by Chester Irving Barnard in his book, 'The Functions of an Executive' in 1938, [5] and since then this terminology has evolved into a discourse of its own. Decision defines the fate and create history, decisions can make a deep impact on the fabric of time and decision can influence a giant to be lost in nonentities in the pages of history. Decision defines the end of the planning process and the beginning of action taking [7]. The decisions are powerful and hence everyone wishes to make the optimal and rational decision, the cost of a wrong decision can devastatingly high hence organizations try their utmost to avoid making wrong decisions.

To be able to make decisions that are correct and rational and offer competitive advantages, the corporate world uses many tools like decision support systems (DSS), expert support systems (ESS), knowledge management systems (KMS), predictive analytics, fuzzy logics, frameworks and many other models. Knowledge management can be defined as an ongoing process of management of knowledge from all sources, internal and external, official and official, explicit and tacit and extracting value from it [11]. Knowledge management is being progressively more used by managers and professionals to be better aware of the situation and making decisions with deeper insights and more knowledge at hand [6, 8, 10, 19, 24, 30].

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 134–145, 2015.
DOI: 10.1007/978-3-319-21009-4_11

2 Literature Review

2.1 Importance and Eminence of Decision Making

Decision making is considered as a vital skill for professional today and enterprises spend a large amount of their budgets in developing systems that can facilitate effective decision making [23, 34, 35]. Management of knowledge, both tacit and explicit is an indispensable part of effective decision making [13, 21, 32]. The right knowledge, the accurate information and in a timely manner is what can help organizations achieve sustainable competitive advantage. Organizations throughout the globe invest billions of dollars to gain knowledge, they develop data warehouses to host archives of knowledge and perform numerous statistical measures and data modelling, mapping and mining techniques to extract value from it and gain a clear picture of the current circumstances and even predict estimates for future [4, 35]. Knowledge management is being increasingly applied by many businesses in the Middle East and North Africa [20]. UAE, which has a vision to diversify its economy also has taken numerous initiatives to better preserve knowledge and leverage it for competitive advantage [12]. Knowledge management is used for many things but one of the uses include making effective, rationale and right decision. As decisions drive businesses, and it is believed that the more information pertaining to a problem one has, the better situational analysis can be done and a more rational and effective decision can be taken by the decision maker [26].

2.2 Organizational Decision Making and UAE

Decision making has always been considered as an important trait in every culture. UAE is considered a melting pot of cultures [29], there exists a rich diversity of cultures and various different styles of management influenced by culture [17]. The managerial styles of administration and decision making also differ and vary significantly, the local Emiratis, other Arabs of the regions, the Asians and the western expatriates all possess a different style of decision making [3]. The importance of making the right decision with complete understanding needs a better situation analysis [33] and to analyze the situation, one need knowledge management to capture and organize all the knowledge. Hence, one can infer that knowledge management can play a very important role in effective decision making [10] and organizations in UAE must take advantage of this fact.

2.3 Impact of Decisions

Decisions can define the fate of any organization. A wrong decision can impact the company for years to come and a single right decision can create a sustainable competitive advantage. Knowledge management can impact the strategic decision making process [24]. Businesses are actively involved in knowledge management and the with use of various ICT based systems like data mining and decision support systems

(DSS) use the knowledge for economic growth [15]. The discipline is hardly 2 decades old, but has been able to create waves. In the simplest form, knowledge management is the process of managing the organizational knowledge, it is an ongoing and continuous process that captures, develops, shares and facilitates the better management of knowledge create by the organization [25].

2.4 Role of Knowledge and Knowledge Economy

Knowledge is considered as power and the present century has also been coined as the century of knowledge, all great nations are striving to develop a knowledge economy, these prove that knowledge management to be a vital for organizations too. Access to knowledge at the right time is critical, it can help the managers make the right decision promptly and swiftly, especially when there is a lot of unneeded data or information overload [18]. Studies have shown that companies lose billions of dollars due to lost knowledge, knowledge management systems aim to capture the knowledge created, not just the explicit knowledge, but also the less quantifiable tacit knowledge, that can serve as a driver for business [22].

2.5 Knowledge Management and Decision Making

Knowledge management has many benefits and is used by almost all industries and segments, be it governmental organizations, private ones or not for profit ones, every organization that values time and knowledge and faces a competition has use knowledge management systems.

2.6 Hypothesis

The following hypothesis were formed;

- H1: Knowledge management can significantly impact the decision making process
- H2: Knowledge management can impact the quality of decisions
- H3: Better situation analysis is directly correlated with better decision making

2.7 Hypothetical Framework

Better decision making is dependent on better understanding of the situation and a good understanding develop when knowledge is managed.

3 Research Methodology

The terminology research methodology refers to the procedures and methods that were used to gather data and analyze it for the purpose of research.

3.1 Research Approach

Researchers can choose among a variety of methods and techniques for the research depending upon the time constraints and the research questions. This research follows a quantitative research approach. A quantitative research requires quantification of variables and application of statistical processes to reach a conclusion.

3.2 Research Design

The research design is causal that is ideally used when one wished to see the impact of one variable on another. The research sample applied was convenience sampling. Further details of the research methodology are listed in Table 1.

Table 1. Research design

Steps of research	Description
Sample selection	A sample of 25 C- level managers were selected that had a significant number of years of experience in corporate organizations and government organizations A convenience sampling technique was used as approaching C-Level executives is a little difficult. Sample for the study from a larger population, according to random starting point to fixed or periodic interval
Population frame	The population frames of the research study are categorized into three frames. These are banks and financial institutions, government enterprises and non-profit
Type of study	As purpose suggest this study related casual study, which interpret that factors which impact on information sharing, which make it successful and successful mentoring effect on the organization. Casual research is focused on finding how one thing can influence others, or in other words, how one variable can cause any change in other variable [16]
Measure and measurement	The measurement scale of the questionnaire is in 5–1. I use Likert scale with scaling from 5 to 1. Where 5 is Strongly Disagree and 1 is Strongly Agree

3.3 Research Sample

The process of research initiated with the identification of the sample frame that was executive and managers, a convenience sampling techniques was used and 50 people were sent a self-administered questionnaire. 25 questionnaires were returned back.

3.4 Research Instrument

The questionnaires were simple, divided into 3 sections. Section 1 consisted of the 10 questions that questions the importance of quality decision making. The Sect. 2 consisted of again 10 questions that about the situational analysis and its role in decision making. Finally the last section consisted of 10 questions covering various aspects of knowledge management including knowledge capturing and knowledge creation, knowledge mapping, knowledge sharing, knowledge transfer, new knowledge generation, knowledge codification and knowledge application. The hypothetical framework developed suggested that effective decision making is dependent on better situation analysis and application of knowledge management.

$$\begin{aligned}\textbf{Effective Decision Making} = {} & \textbf{Better Situation Analysis} + \\ & \textbf{Application of Knowledge Management}\end{aligned}$$

3.5 Statistical Tools and Techniques

The aim of the research is to explore the impact of knowledge management on the quality of decisions. The research questions under discussion included;

1. Does knowledge management improve the quality of decision making
2. Can knowledge management offer a better insight to business and more clear situation analysis
3. Knowledge management system can improve help managers make informed decisions
4. Better decisions are a result of better knowledge management.

Considering the research aim and objective the following tests were conducted on the data gathered;

1. Reliability analysis
2. Correlation analysis
3. ANOVA or analysis of variance
4. Regression analysis

3.5.1 Research Equation

$$\begin{aligned}\textbf{Effective Decision Making} = {} & \textbf{Better Situation Analysis} + \\ & \textbf{Application of Knowledge Management}\end{aligned}$$

This equation will be tested to determine whether our hypothesis stand true or not using the regression model.

3.5.2 Dependent and Independent Variables

As stated earlier as well that is research is a causal research that is, it aims to investigate the impact of certain variables on one another. In our research model, the ability to make effective decisions is the dependent variables which is depending on two independent variables the better situation analysis ad and the application of knowledge management.

4 Results, Limitations and Recommendations

4.1 Descriptive Analysis

Descriptive analysis is a representation of how the variables are demographically divided. The sample that we had selected could be divided as per numerous factors like;

- Age bracket
- Gender
- Education acquired
- Years of professional experience.

Table 2. Demographic analysis

Demographic variables	Gender	Age		Education level		Professional experience	
Male	16 (64 %)						
Female	9 (36 %)						
21–25		0	(0 %)				
26–30		1	(4 %)				
31–35		3	(12 %)				
36–45		15	(60 %)				
46–55		4	(16 %)				
>55		3	(12 %)				
High school				0	(0 %)		
Diploma				0	(0 %)		
Graduate				5	(20 %)		
Post Graduate				10	(40 %)		
Ph.D.				10	(40 %)		
0–3 years						0	0
4–5 years						0	(0 %)
6–10 years						1	(4 %)
10–15 years						5	(20 %)
16–20 years						13	(52 %)
20–30 years						3	(12 %)
>30 years						3	(12 %)

From the Table 2, we can clearly evaluate that the sample consisted of a well blend of both male and females, there were 2 thirds of male participant and one third female participant. The majority of the sample was of more than 35 years of age. Nearly 40% of the sample had done post-graduation and 40% were Ph.D. Scholars. 52% of the sample had 16–20 years of work experience, 20 had 10–15 and 12% had 20–30 and more than 30 years of professional experience.

4.2 Reliability Analysis

Reliability analysis or Cronbach is not a statistical test but it is an assessment of the research instrument or the instrument that is used for the data collection purposes. We had opted for a self-administered questionnaire which had 30 items and later 3 more items were computed from the existing item only (Table 3).

Table 3. Reliability statistics

Cronbach's alpha	N of items
.829	33

Usually it is observed that the reliability of instruments with lesser sample sizes is low, but in our case it appeared to be 82.9%, which is a healthy level and one can conclude the research instrument as reliable. It will produce consistent results again and again.

4.3 Correlation Analysis

Correlation is the association or relationship among variables. As it was stated previously as well that this is a causal research, hence we need to see how much association variables has with another.

The results of the Table 4 indicate that situational analysis is positive and strongly bonded with decision making. They have 80% association any change in one of them will influence the other also. The association and bonding between knowledge management and decision making was relatively weak and insignificant at 28.9%.

4.4 ANOVA and Regression Analysis

"R" in the Table 5 is a representation of correlation, the overall correlation between the 3 variables is 88.4%.

ANOVA is an analysis of variance. ANOVA doesn't consider any individual value but the average value of a group of variables (Tables 6).

The beta or "B" in the Table 7 is the constant that shall hold true in all circumstances as per the hypothetical framework and research equation. The levels of significance for the situation analysis and the knowledge management are less than the p-value hence we can claim that both are significant for effective decision making.

Table 4. Correlations

		DM	SitAnal	KM
DM	Pearson correlation	1	.801**	−.086
	Sig. (2-tailed)		.000	.690
	N	25	25	24
SitAnal	Pearson correlation	.801**	1	.289
	Sig. (2-tailed)	.000		.171
	N	25	25	24
KM	Pearson Correlation	−.086	.289	1
	Sig. (2-tailed)	.690	.171	
	N	24	24	24

**Correlation is significant at the 0.01 level (2-tailed)

Table 5. ANOVA and regression model summary

Model	R	R square	Adjusted R square	Std. error of the estimate
1	.884[a]	.782	.761	.16607

a. Predictors: (Constant), KM, SitAnal

Table 6. Analysis of Variance ANOVA[a]

Model		Sum of squares	df	Mean square	F	Sig.
1	Regression	2.077	2	1.039	37.655	.000[b]
	Residual	.579	21	.028		
	Total	2.656	23			

a. Dependent variable: DM
b. Predictors: (Constant), KM, SitAnal

Table 7. Regression analysis coefficients[a]

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. error	Beta		
1	(Constant)	1.289	.489		2.636	.015
	SitAnal	.988	.114	.919	8.637	.000
	KM	−.298	.090	−.351	−3.301	.003

a. Dependent variable: DM

5 Discussion and Outcome

There is a popular phrase in English that, "a little knowledge is dangerous". The phrase can also be applied to the present world of business. The business and the socio-political landscape is changing rapidly. The process of making decision is also becoming complex by the day and managers and business leaders are under immense pressure to make accurate predictions and vital decisions to maintain the competitive advantage of their businesses [28]. Can the application of knowledge management facilitate the decision making process? Do leaders use and refer to knowledge management systems for making decisions? Can the application of knowledge management help leaders and top management is conducting a better and clearer situation analysis? To seek answers for these questions, this research study was under taken. It is evident from the study of the secondary literature that both decision making and knowledge management are important managerial functions and in the recent years managers have resorted to the use of knowledge management to facilitate the decision making process [1, 12, 24]. The world of business is the world of decisions, here the ability to make good, rationale and logical decisions is considered as an asset. In the complex world of business people can't just rely on their hunch or gut feeling, there has to be evidence backed practices or best practices. The process of organizational decision making has its own dynamics and nomenclature, the process itself has a significance and has turned into a subject of academic debate [27].

Strategic decision making is a critical function for members of the C-suite and any wrong decision can create a dent in the progress of the company for years to come hence the top management, the board of directors, the member of the C-suite are very concerned about the decision making process and wish to acquire a complete situation analysis before making any decision [2, 14, 31]. CEO's are under immense pressure to make rational and logical decision and hence many members of C-Suite use various insights and deeper discernments and precise situational analysis before making any decision. Knowledge is considered as power and the present century has also been coined as the century of knowledge, all great nations are striving to develop a knowledge economy, these prove that knowledge management to be a vital for organizations too. Access to knowledge at the right time is critical, it can help the managers make the right decision promptly and swiftly, especially when there is a lot of unneeded data or information overload [18]. Studies have shown that companies lose billions of dollars due to lost knowledge, knowledge management systems aim to capture the knowledge created, not just the explicit knowledge, but also the less quantifiable tacit knowledge, that can serve as a driver for business [22].

Mangers feel anxious and also uneasy when they have to make a decision that can have grave consequences. The question under discussion is how a simple system can help in decision making, as discussed, that a knowledge management system is a specialized system that is responsible for creation of knowledge, capturing of knowledge, sharing of knowledge and application of knowledge from all sources be it internal or external. KMS are backed with expert support systems or/and decision support systems as well, they gather information generated from multiple sources, code them and classify them there forth when information is needed a complete picture is offered by the knowledge management system which is more richer and deeper than other ICT based systems that are commonly deployed. KMS also has predictive analytics hence managers can review the impact of their decision. KMS improve the quality of data and hence a better situational analysis is created. A clear situation analysis of both internal and external, micro and macro environments at the same time offer managers with the necessary intelligence needed to make strategic decisions. From the statistical analysis, we can conclude that knowledge management and situational analysis are both important but situation analysis has a better and stronger positive association with effective decision making. Our study has 3 hypothesis;

- H1: Knowledge management can significantly impact the decision making process
- H2: Knowledge management can impact the quality of decision
- H3: Better situation analysis is directly correlated with better decision making

From data collected from C-Level executive, where the pinnacle of decision making is performed, we can conclude that all three hypothesis are valid and acknowledge. KMS are the need of the time and organizations must adopt them not only for making major decision but also internal decisions. Knowledge management and other ICT based software are helping mangers to make logical and rational decisions but they rely more on situational analysis. The rational tools that offer a deeper insight of the market, the company the internal and external forces of change are considered by many when making critical decisions.

Knowledge management is highly used in the banking sector and other financial services originations, they use knowledge management systems for big data analysis and making investment decisions, it is high time that even SME's and entrepreneurs implement some sort of knowledge management processes if not a comprehensive and business intelligence loaded KMS. In the knowledge economy, knowledge is the currency and decisions are the actions that define the fate of corporations.

6 Conclusion and Recommendations for Future

We know that knowledge management has been doing wonders for the business world. Organizations that exploit and leverage knowledge are better able to sustain and generate a sustainable competitive advantage. The study which was aimed at finding the impact of knowledge management and situational analysis on the decision making skills of organizations and individuals found that it is use. There is a significant association between decision making, knowledge management and decision making. The use of knowledge management software and systems help managers and organizations capture knowledge that is usually lost or unable to be formally documented. When an employee quits the organization, his/her replacement doesn't have the learning curve nor is immediately that productive. Knowledge management systems can help in avoiding any losses.

Scholars studying the domain can also conduct researches on the same topic but also including personal or cultural behavioral traits. As UAE is a culturally rich and diverse society, the expatriate population makes a significant chunk of the population belong to different cultures and nationalities. The personal styles can also impact their ability to make decisions, hence does the use of knowledge management reduce the influence of culture or nationality and help in formation of a uniform decision making formula can be studied. Another dimension for future would be the comparison of using a comprehensive KMS versus an operational ERP or data mine and is the costs associated with KMS justified in terms of the empowerment and insights it offers to the managers for rational decision making or not. UAE or Middle East specific researches would also offer new light on the domain, a comparative analysis of knowledge management in a western country and an Arabic country within similar organization can offer a new dimension to the already existing knowledge and also open new avenues for improvement of the present KMS.

References

1. Alavi, M., Leidner, D.E.: Review: knowledge management and knowledge management systems: conceptual foundations and research issues. MIS Q. **25**(1), 107–136 (2001)
2. Albright, S., Winston, W., Zappe, C.: Data Analysis and Decision Making. Cengage Learning, Boston (2010)
3. Ali, A.J., Azim, A.A., Krishnan, Krish S.: Expatriates and host country nationals: managerial values and decision styles. Leadersh. Organ. Dev. J. **6**(6), 27–34 (1995)
4. Argote, L.: Organizational Learning: Creating, Retaining and Transferring Knowledge. Springer Science & Business Media, Berlin (2012)
5. Barnard, C.I.: The Functions of the Executive. Harvard University Press, Cambridge (1968)
6. Bolloju, N., Khalifa, M., Turban, E.: Integrating knowledge management into enterprise environments for the next generation decision support. Decis. Support Syst. **33**(2), 163–176 (2002)
7. Buchanan, L., O'Connell, A.: A Brief History of Decision Making. Harvard Business Review, Cambridge (2006)
8. Choo, C.: The knowing organization: how organizations use information to construct meaning, create knowledge and make decisions. Int. J. Inf. Manag. **16**(5), 329–340 (1996)
9. Clarke, H.C.L.: Deciding How to Decide. Harvard Business Review, Cambridge (2013)
10. Courtney, J.F.: Decision making and knowledge management in inquiring organizations: toward a new decision-making paradigm for DSS. Decis. Support Syst. **31**(1), 17–38 (2001)
11. Demarest, M.: Understanding knowledge management. Long Range Plan. **30**(3), 321–322–374–384 (1997)
12. Dubai School of Governance: UAE in the context of knowledge management (2014). http://www.dsg.gov.ae/SiteCollectionImages/Content/DeG%20Documents/April-2014-en.pdf. Accessed 27 March 2015
13. Grant, R.M.: Toward a knowledge-based theory of the firm. Strateg. Manag. J. **17**(2), 109–122 (2014)
14. Heller, F.A.: Decision-Making and Leadership. Cambridge University Press, Cambridge (1992)
15. Hossain, N., Sarkani, S., Mazzuchi, Thomas A.: The integration and impact of effective knowledge management systems in the advancement of the decision making process in complex organizations to realize economic growth. Procedia Comput. Sci. **8**, 100–105 (2012)
16. Hair, J.F., Wolfinbarger, M., Money, A.H., Samouel, P., Page, M.J.: Essentials of Business Research Methods, 1st edn. M.E. Sharpe, Armonk (2011)
17. Kumaravadivelu, B.: Cultural Globalization and Language Education. Yale University Press, New Haven (2008)
18. Lisa, Q.: Why knowledge management is important to the success of your company (2012). http://www.forbes.com/sites/lisaquast/2012/08/20/why-knowledge-management-is-important-to-the-success-of-your-company/. Accessed 28 March 2015
19. Lynam, T., De Jong, W., Sheil, D., Kusumanto, T., Evans, K.: A review of tools for incorporating community knowledge, preferences, and values into decision making in natural resources management. Ecol. Soc. **12**(1), 5 (2007)

20. Magd, H., McCoy, M.: Knowledge management in the middle east the leadership challenge. In: Soliman, F. (ed.) From Knowledge Management to Learning Organisation to Innovation: The Way Ahead!, pp. 251–300. Cambridge Scholars Publishing, Cambridge (2015)
21. Mårtensson, M.: A critical review of knowledge management as a management tool. J. Knowl. Manag. **4**(3), 204–216 (2000)
22. Mind Tools: Knowledge management: making the most of intellectual assets (2012). http://www.mindtools.com/pages/article/newISS_87.htm. Accessed 29 March 2015
23. Mitchellette, R.J.: Entrepreneurial Decision Making: A Must-read for Every Aspiring Entrepreneur. Xlibris Corporation, Bloomington (2008)
24. Nicolas, R.: Knowledge management impacts on decision making process. J. Knowl. Manag. **8**(1), 20–31 (2004)
25. Nonaka, I.: Perspective—tacit knowledge and knowledge conversion: controversy and advancement in organizational knowledge creation theory (2009). http://pubsonline.informs.org/doi/abs/10.1287/orsc.1080.0412. Accessed 27 March 2015
26. Onwubiko, C.: Situational Awareness in Computer Network Defense: Principles, Methods and Applications: Principles, Methods and Applications. IGI Global, Hershey (2012)
27. Pettigrew, A.M.: The Politics of Organizational Decision-Making. Routledge, New York (2014)
28. Proctor, T.: Creative Problem Solving for Managers: Developing Skills for Decision Making and Innovation. Routledge, New York (2014)
29. Reiser, D.: Analysis of Cultural Differences in Dubai. BWV Verlag, Berlin (2010)
30. Rubenstein-Montano, B., et al.: A systems thinking framework for knowledge management. Decis. Support Syst. **31**(1), 5–16 (2001)
31. Shapira, Z.: Organizational Decision Making. Cambridge University Press, Cambridge (2002)
32. Smith, E.A.: The role of tacit and explicit knowledge in the workplace. J. Knowl. Manag. **5** (4), 311–321 (2001)
33. Stone, D.: Policy Paradox: The Art of Political Decision Making. W.W. Norton & Company, New York (2012)
34. Vaidya, K.: Inter-Organizational Information Systems and Business Management: Theories for Researchers: Theories for Researchers, 1st edn. IGI Global, Hershey (2011)
35. Vercellis, C.: Business Intelligence: Data Mining and Optimization for Decision Making, 1st edn. Wiley, New York (2011)

Measurement Model of Relationship Between Knowledge Management Enablers and Knowledge Management Performance in Public Sector

Karimah@Aminah Abu Bakar[1], Nor Zairah Ab Rahim[2(✉)], and Roslina Ibrahim[2]

[1] Research Management Division, Malaysian Administrative Modernisation and Planning Unit (MAMPU), Prime Minister's Department, Federal Government of Malaysia, 62502 Putrajaya, Malaysia
karimah@mampu.gov.my

[2] Advanced Informatics School (AIS), Universiti Teknologi Malaysia (UTM), 54100 Kuala Lumpur, Malaysia
{nzairah, iroslina.kl}@utm.my

Abstract. This study investigates the relationship between KM enablers and the KM performance in the Malaysian public sector. Five main independent variables of KM enablers; organizational culture, organizational structure, technology, people and political influence were tested against KM performance using structural equation modeling (SEM). Five hypotheses were tested on the relationship between KM enablers and KM performance. The sampling population of this study was Ministry of Natural Resources and Environment (NRE), Malaysia. The result confirmed that organizational culture, organizational structure, technology and people are the distinct constructs for KM enablers. The hypothesis testing indicates that there is a significant relationship between KM enablers and KM performance. The result indicates that Malaysian public sector should cultivate on KM enablers in order to implement effective KM.

Keywords: KM performance · KM enablers · Public sector

1 Introduction

Knowledge is known as an intangible yet an important asset for gaining continuous competitive advantage [1]. According to Natarajan and Shekhar [2], knowledge does not merely represent power but also wealth. Organisations that successfully manage their knowledge have a greater ability to act and adapt, and are thus better poised to thrive in this complex new business environment [3]. In other words, improving KM will ultimately lead to organizational improvements.

With such compelling rationale for KM in organisations, the leadership and management of an organisation should take the knowledge management issues seriously. A good starting point would be to understand their current level of KM performance. Performance measurement is crucial in KM as it serves as the foundation that enables an

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 146–159, 2015.
DOI: 10.1007/978-3-319-21009-4_12

organization to evaluate, control, and improves its knowledge processes [4]. However, relatively little study has been performed on KM performance in organizations especially in the public sectors and even less in the developing countries [5].

Today, public sector firms are known as knowledge based organizations since knowledge is their most vital asset [6]. As knowledge is a central resource of the public sector service, effective KM is a significant public management challenge for providing excellent public sector service to constituencies at all levels. In Malaysia, the government's concern in developing countries through the knowledge economy has become clearer. The Malaysian Administrative Modernisation and Management Planning Unit (MAMPU) has come out with a KM blueprint in 2011 which outline the strategies and recommendations for setting KM initiatives within the public sector environment [7]. Prior to the development of the blueprint and till recently, many Malaysian public organizations have embedded KM initiatives within their organizations.

Despite the active commitment towards better KM, organizations are struggling with their KM implementations. Organizations suffer from difficulties in evaluating performance of their KM initiatives due to the lack of a proper framework for assessing the current status of KM which later create doubt over the basic concept itself. To address these problems, this study proposes a framework to assess the current status of KM in the public sector organization through identifying the relationships of KM enablers towards the KM performance.

2 The Research Model

This research aims to investigate the implementation of KM in the Malaysian Public Sector by examining the relationship between KM enablers and KM performance in the organization. Based on the framework by Syed-Ikhsan and Rowland [5], authors propose five major factors as enablers of KM. The five factors are the organizational culture, the organizational structure, the technology, the people and the political

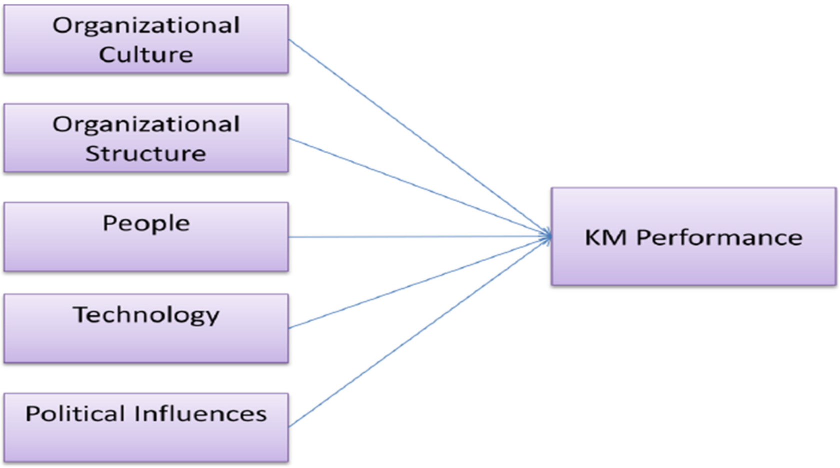


Fig. 1. Conceptual framework

influences. These factors could be measured to explore the relationship between specific variables and performance of KM in public organizations. The conceptual framework is then established as shown in Fig. 1.

3 KM Performance

Performance measurement is crucial in KM as it serves as the foundation that enables an organization to evaluate, control, and improves its knowledge processes. Improving KM will ultimately lead to organizational improvements Kuah and Wang [1].

According to Wei et al. [15], KM Performance measurement refers to the assessment methods of KM and its relationships with organizational performance. Many companies fail to measure their KM efforts due to reasons such as failure to operationally define performance, failure to relate performance to the process, failure to define the boundaries of the process, and misunderstood or misused measures. Thus, it is clearly necessary to include a performance measurement system as a key factor to successful KM implementation.

This study adopted the dimensions used from previous studies for measuring KM. There are five dimensions constructed in previous studies that show the reliability of the dimensions on KM performance. The five dimensions are speed of knowledge transfer, accuracy of knowledge transfer, reliability of knowledge transfer, tacit knowledge and explicit knowledge [5, 8, 9]. These dimensions are the essentials of KM performance and proven to be reliable as shown in previous studies.

3.1 Speed of Knowledge Transfer

The knowledge transfer is a spiral process as described in SECI Model by Nonaka et al. [16] through its four modes of knowledge conversion and the evolving spiral movement: Socialisation, Externalisation, Combination, and Internalisation process. How fast knowledge can be transferred is the most important element that needs to be identified. According to Bloodgood and Salisbury [10], knowledge transfer may “lead to advantage through speedier deployment of knowledge to portions of the organizations that can benefit most by it”. Davenport and Prusak [11] refer to speed as velocity in which knowledge moves through an organization. With a proper infrastructure, individuals in the organization are likely to obtain information faster and can make faster decisions [5].

3.2 Accuracy of Knowledge Transfer

However, speed alone will never solve the problem unless knowledge is transferred accurately. The most important aspect of knowledge management in the organizational setting is for the knowledge to be transferred to the right locations, to the right person, at the right time and for the right reasons [5]. In other words, if information and knowledge can be transferred faster but reach the wrong person, it will cause more problems to any organization. Explicit knowledge such as patents, instruction manuals, written procedures, best practices, lessons learned and research findings are normally and should be shared with a high degree of accuracy [12].

3.3 Reliability of Knowledge Transfer

The third dimension that also needs to be considered is the reliability of the knowledge in the organization. Reliability of data, information and knowledge assets in the public sector organization is very important as inaccurate knowledge might cause problems to the government as a whole [5]. Chong et al. [8] use reliability of knowledge transfer to measure the transfer process of explicit knowledge.

3.4 Knowledge Assets

The creation of knowledge assets are very important to all organizations as they play a major role in all decision making. However, very little attention is given to how knowledge is created and how the knowledge creation process can be managed. According to Bloodgood and Salisbury [10], every organization needs to identify where knowledge resides in the organization. With reliable collections of knowledge assets, then knowledge can be transferred to the respective person at the right time and at the right place with great accuracy. In other words, the performance of knowledge transfer depends more on the availability and the accessibility of the knowledge assets.

4 KM Enablers

KM enablers are critical factors that need to be developed in an organization that serve as driving forces to solidify knowledge management [13]. Each of the selected KM enablers will be explained further in the following sections.

4.1 Organizational Culture

Culture is regarded as the key factor since it determines the effects of other variables such as technology and management techniques on the success of KM [14]. Wei et al. [15] define organizational culture as the general knowledge sharing climate of an organization as related to an integrated pattern of human behaviors – including thoughts, speeches, actions, and artefacts. The authors further argue that a company's ability to use knowledge therefore depends primarily on the level of enthusiasm among employees that is ultimately influenced by the organization's culture.

However, according to Cong and Pandya [12], getting an organisation's culture 'right' for KM is typically the most important and yet often the most difficult challenge. For the purpose of this study, organizational culture is manifested through two dimensions that is sharing culture and individualism [5], which will be further explained in the following subsections.

Sharing Culture. According to Chong et al. [8], sharing culture will not occur in an organization unless its employees and workgroups display a high level of trust and co-operative behaviour. Change in culture and individual behaviour must aim towards encouraging the use of knowledge not for individual advantage but for the benefits of the organization as a whole.

Erwee et al. [14] suggests that organizational culture is an essential enabler for an effective knowledge culture with trust identified as the single most powerful cultural dimension in this process. Skadiang [17] explained that KM is basically a cultural construct because different cultures have different ways of structuring meaning, for example in the USA, the adage 'knowledge is power' is relevant whereas in Japan, knowledge is only knowledge when it is shared as one's status depends on how much one contributes to the group. Any changes need to be developed must be in line with the existing organizational culture.

Individualism. Cong and Pandya [12] argue that knowledge sharing is not a natural act in public organisations. This is due to the fact that the structure of the public sector organisations has traditionally been compartmentalised. 'Need to know' basis is part of public sector culture. They further pointed that 'knowledge is power', 'what's in it for me', and 'not invented here' syndrome are typical mindsets of the manager and staff in public organizations. In such an environment, information and knowledge are hardly ever shared across different units and different organisational levels. However, people share knowledge for some reasons such as reciprocity, reputation and prestige, or sometimes just for altruistic reasons. Therefore, a knowledge sharing culture need to be created in order to change the attitude and behaviour of the people and reduce barriers because there can be no KM without knowledge sharing [13].

Therefore, it is hypothesized that:

H1: Organizational culture has a significant relationship with the KM performance.

4.2 Organizational Structure

Organizational structure is "the structure of an organization is what follows from a division of work, tasks and responsibilities, both horizontally and vertically. Knowledge sharing is likely to occur within a larger group of individuals in more decentralized organizations. In addition, matrix structures and an emphasis on leadership instead of management also facilitate greater knowledge sharing primarily by cutting across traditional departmental boundaries" [15]. Organisational structure also refers to the way employees are organised into teams (informal and formal), and interact within teams, the set of roles and goals of each team, and how it is being related to organisational strategy [18]. The organisational culture is considered to be a critical factor in building and reinforcing knowledge creation and knowledge management in organization as it impacts how members learn, acquire, and share knowledge [19].

For the purpose of this study, the organisational structure will only be discussed in terms of its influence on the communication flows between departments and proper documentation of policies, procedures and regulations imposed in the Ministry and how they help to create and transfer knowledge [5].

Document Confidentiality Status. Syed-Ikhsan and Rowland [5] argue that one of the factors that influences the creation and transfer of knowledge in an organization is the status of information and documents. Certain items of information and documents are restricted to certain levels of employee, which prevents the flow of knowledge across the organization.

Communication Flows. In a traditional model, large organizations normally have many layers of managers where "formal reporting structures are more detailed at the top than the bottom" [11]. Decision making flows vertically up and down this chain of command, and often communication also flows only up and down this chain of command. Monavvarian and Kasaei [18] claim that most of the communication functions are "top-down" and therefore too slow to meet employee needs. Consequently, it takes too much time for information to filter down through every level of the organization. According to the authors, effective top-down and bottom-up communication is very important in making existing knowledge profitable to the organization. If an organization supports communication networks that operate freely, where knowledge providers and knowledge seekers can access information and knowledge through the shortest path, it will certainly enhance knowledge creation and knowledge transfer in the organization [5]. It is hereby anticipated that: H2: Organizational structure has a significant relationship with the KM performance.

4.3 Technology

Researchers and practitioners who subscribe to the IT-dominant school of KM thought believe that if extensive computer networks and communications tools for group collaboration are developed, people will have a greater propensity to share information and knowledge [20]. The use of powerful IT and communications tools can support organisations but in a dynamic environment where businesses face unpredictable and unique problems, IT is at best an enabler since it is only through people that information can be interpreted and transformed into knowledge [21].

ICT Infrastructure. According to Wei et al. [15], technological infrastructure refers to the infrastructure of tools, systems, platforms and automated solutions that centralize and enhance the development, application, and distribution of organizational knowledge. Syed-Ikhsan and Rowland [5] find that effective KM depends on people sharing their knowledge through computer facilities that users throughout the organization have access to. In other words, up-to-date ICT infrastructure will help employees to create, share and transfer knowledge within the organization.

ICT Tools. Chong et al. [8] claims that the ICT infrastructure should promote the efficient and effective capture of both tacit and explicit knowledge and to support knowledge sharing in the entire organization. In fact, effective KM depends on people sharing their knowledge through computer facilities together with the users of knowledge throughout the organization being able to have access to the organization's knowledge base. Communication networks, electronic mails, intranet, data warehousing, and decision support systems are some of the basic elements of KM technology infrastructure [8].

ICT Know-How. Know-how refers to skills, such as the capability to do something. It might relate to the skills of manual workers [22]. Chong et al. [8] claim that the more training provided for information and communications technology (ICT) skill upgrading, the more knowledgeable the individuals would be in utilizing the ICT tools and KM technologies. Hence, more knowledge can be transferred and shared within and outside the organization. Therefore, it is posited that: H3: Technology has a positive relationship with the KM performance.

4.4 People

Another important construct in knowledge management is people as the brain of the people in organizations have an unlimited capacity for tacit or explicit information and knowledge [22]. Knowledge comes as a person uses information and combines it with their personal experience. Much of the knowledge one acquires and gathers in one's head has its own value, and it is that which makes each of us unique and valuable to society as a whole and to organisations. One of the main challenges, in creating knowledge based organizations in Malaysia is, changing people's behaviour and retaining talented people [23]. These issues are much related to human resource (HR) policies in organization. That explains why this construct is grouped together as 'people/human resources' by Syed-Ikhsan and Rowland [5]. Strategic HR policies will also influence how organizations manage knowledge as it could motivate employees' willingness to acquire, share, and apply knowledge within organizations [24]. Therefore, for the purpose of this study, the dimensions of people will be discussed in terms of posting, training and staff turnover [5] which also much related to HR policies.

Posting. One of the main criteria that will be looked into this area is the posting of officer to particular positions, undertaken by the Public Service Department (PSD) and other related agencies to the Ministry of Natural Resources and Environment. According to Chong et al. [8], employees bring to an organization their prior education, experience, knowledge and skills and will add value to the organization. This aspect is very important, as knowledge is likely to be created easily if employees are placed in the right positions.

Training. Knowledge gained by employees through job training will enable them to translate their knowledge into the organization's routines, competencies, job descriptions, and business processes, plans, strategies, and cultures [8]. While the human brain has the unlimited power that is beneficial to the organization, the right skills for creativity must be tapped [22]. This suggests that employees should be provided with the right and continuous training in order to enrich their knowledge and improve their capabilities. Syed-Ikhsan and Rowland [5] claim that employees with a lack of adequate training, or explicit knowledge, will struggle to keep up. Therefore it is important for the organization to have a proper training program to enable employees to gain knowledge and contribute to the creation and transfer of knowledge in the organization.

Staff Turnover. Syed-Ikhsan and Rowland [5] find that the problem of staff turnover also happens in all public organisations. Employees leaving the civil service pose a challenge to knowledge initiatives, because organizational knowledge assets may be lost as people retire or leave for other positions. Therefore, it is necessary to have an appropriate procedure to ensure that information and knowledge can be kept in the organization. The following hypothesis thus ensues: H4: People has a positive relationship with the KM performance.

4.5 Political Influences

In a public organization, political influences have a great impact on the creation of knowledge assets. Sometimes there are unwritten policies or directions that need to be followed. The assumption is political influence in a public organization has an impact to the effectiveness of knowledge transfer [5]. It is postulated that:

H5: Political influences has a positive relationship with the KM performance.

5 Methodology

Ministry of Natural Resource (NRE), one of the public sectors in Malaysia has been chosen for the case study. NRE is responsible in managing the natural resources and environment in Malaysia. Since professional intellects in NRE are known as the primary source of knowledge in natural resources and environment, the future implementation of KM at the NRE is beneficial to leverage on the highly tacit experiences, ideas and expertise of its professional intellects in improving the performance of the Ministry.

5.1 Survey

Data was gathered from the middle management level of the organization. This is due to the fact that they are primarily in the operational and strategic decision making process. Therefore, their responses to the issues raised in the questionnaire can render a high level of credibility and can have long-term consequences for the future success of KM implementation in the NRE. There were 223 middle managers in the organization and the self-administered survey questionnaires were distributed through electronic mail to all of them. 147 questionnaires were returned with 66 % of response rate.

The questionnaire is divided into two sections. Section A consists of questions (items) eliciting demographic and personal background information. Section B consists of 16 questions (items) designed to ascertain general views towards the topic of the study. A five point Likert scale where 1 represents “strongly disagree” and 5 represents “strongly agree” is used in the questionnaire. In this study, the questionnaire was mainly adopted from Syed-Ikhsan and Rowland [5] and Chong et al. [8] who particularly researched on KM in Malaysia public sector and several other researchers researching in the same area elsewhere as shown in the following Table 1.

Table 1. Source of items in constructs

Constructs	Questions no.	No. of item	References
Organizational culture	16–17	6	[5, 8, 13–15]
Organizational structure	18–19	5	[5, 8, 13–15]
Technology	20–22	10	[5, 8, 13–15]
Human resource/people	23–25	7	[5, 8, 13–15]
Political influences	26	2	[5]
KM performance	11–15	13	[5]
Total items		43	

To give respondents a consistent understanding of what KM is, the definition of knowledge and KM will be included in the questionnaire. In this study, KM is defined as a systematic and organised attempt to use knowledge within the organisation to provide services to the public and to improve performance. Knowledge in the organisation includes explicit and tacit knowledge.

5.2 Data Analysis and Results

A pilot test of for the study had been conducted to test the reliability of the dimension used and the data. The pilot test has shown that all respondents understood the questionnaire sufficiently. Cronbach's alpha was calculated for the total of six dimensions consists of 43 items. Result for the pilot test are within 0.634 to 0.880 indicate that the items have acceptable internal consistency for each construct. Therefore, no items were deleted in the actual study.

The actual data analysis on KM enablers and KM performance were conducted with exploratory factor analysis (EFA), confirmatory factor analysis (CFA) and structural equation modelling (SEM). The first step is to assess the suitability of the data for factor analysis. This involves inspecting the correlation matrix for coefficients 0.5 and above, and calculating the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) and Bartlett's Test of Sphericity. The value of KMO is 0.6 and above for both KM enablers and KM performance, and the value of Bartlett's Test of Sphericity is significant ($p = .000$), therefore factor analysis are appropriate.

Once the numbers of factors have been determined, the next step is to interpret them. In this process, the factors are 'rotated' using Varimax rotation. The item is accepted if the loading factor is above threshold of 0.5. Therefore, the result suggested that 2 items from technology dimension and 1 item for people in the KM Enablers constructs should be dropped as their loading factors are below 0.5. While for KM Performance construct, all items are accepted as their loading factors are above 0.5.

Structural equation modeling (SEM) is employed for confirmatory factor analysis (CFA). The purpose of conducting CFA is to reconfirm the result of EFA of the same data. Basically in SEM, there are two models involved in the analysis namely; measurement model, the model that demonstrates the relationship between response items and their underlying latent constructs and structural model, the model that demonstrates the correlational or causal dependencies among the measurement models in the study. SEM also has the ability to assess the unidimensionality, validity and reliability of a measurement model. Unidimensionality is achieved when the measuring items have acceptable loading factors for the respective latent construct. According to Awang [25], in order to ensure unidimensionality of a measurement model, any item with a low loading factor low than 0.5 should be deleted. Both items that measured the political influences indicated factor loading below the threshold of 0.5, therefore both items including the construct were deleted which resulted in the dropped of the construct from the model. After running the CFA analysis, 24 out of the 43 items were deleted. At the end of the analysis, the validity and reliability of the measurement model have been achieved since all the index value for each category is within the level of acceptance. Table 2 summarizes the analysis.

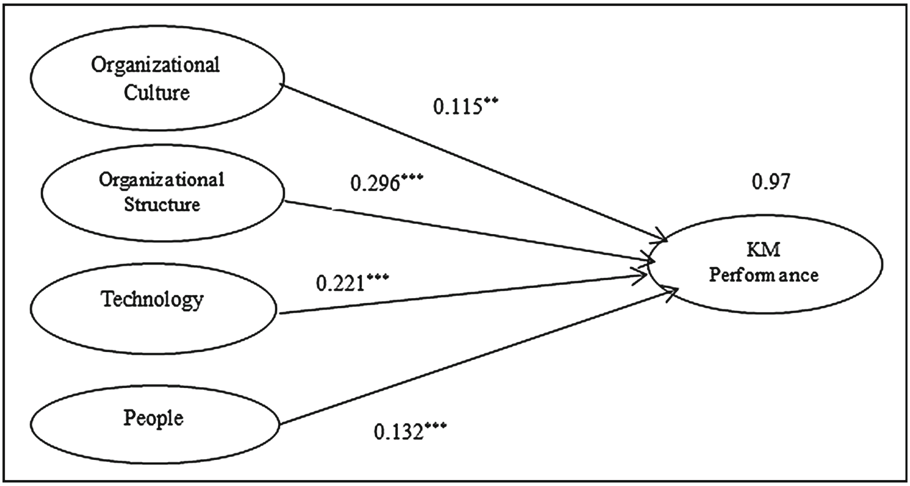
Table 2. List of final items

Construct	Items	Factor Loading	Remarks
Organizational culture	OC5	.983	4 out of 6 items deleted
	OC6	.836	
Organizational structure	OS1	.925	2 out of 5 items deleted
	OS2	.882	
	OS4	.526	
Technology	T2	.925	5 out of 10 items deleted
	T3	.890	
	T5	.719	
	T6	.624	
	T7	.577	
People	P1	.836	4 out of 7 items deleted
	P2	.974	
	P3	.938	
KM performance	Sp1	.585	8 out of 13 items deleted
	Sp2	.934	
	Sp3	.385	
	Sp4	.963	
	Re2	.555	
	Ac2	.656	

***$p < 0.001$, **$p < 0.01$

Finally, the value of R2 is 0.97 (Fig. 2), which indicate the contribution of the rest of the constructs organizational culture, the organizational structure, the technology, the people in estimating KM performance is 97 %. Therefore, the measure of correlation indicates that the strength of the relationship between the constructs is strong.

The hypothesis testing shows the satisfying estimate value of 0.115 for relationship between KM performance and organizational culture. It indicates that organizational culture has a positive relationship with KM performance in Malaysian public sector. The result is consistent with previous researches by Cong and Pandya [12], Erwee et al. [14], Wei et al. [15], Syed-Ikhsan and Rowland, [5], Chong et al. [8], Skadiang [17] and Yusof et al. [13]. The second hypothesis is to determine the significance between KM performance and organizational structure. The result shows the estimate value of 0.296 for the relationship between KM performance and organizational culture. The result is consistent with previous researches by Wei et al. [15], Monavvarian and Kasaei [18], Rai [19] and Syed-Ikhsan and Rowland [5]. The third hypothesis is to determine the significance between KM performance and technology. The result shows the estimate value of 0.221 for the relationship between KM performance and technology. The result is consistent with previous researches by Bollinger and Smith [20], Bhatt [21], Wei et al. [15], Syed-Ikhsan and Rowland [5] and Chong et al. [8]. The fourth hypothesis is to determine the significance between KM performance and people. The result shows the estimate value of 0.132 for the relationship between KM



***p<0.001, **p<0.01

Fig. 2. SEM for KM enablers and KM performance

performance and people. The result tally with previous researches by Chong et al. [8] and Syed-Ikhsan and Rowland [5].

6 Discussion and Conclusion

This paper highlighted the importance of identifying the KM enablers that could influence the KM performance in public organizations. The KM enablers have been categorized in 4 dimensions: organizational culture, organizational structure, technology and people. These KM enablers are proposed to be measured against the KM performance measures which include: speed of knowledge transfer, accuracy of knowledge transfer, reliability of knowledge transfer, tacit knowledge and explicit knowledge. The measurement of the relationship of the two sets of dimensions formed a conceptual framework that could be used to measure the KM performance in public organizations.

The analysis results reveal that the speed, accuracy and reliability of knowledge transfer in an organization has a direct influence on the KM performance. The findings suggest that public organization need to manage the performance of knowledge transfer in ensuring that the organization can take full advantage of the KM implementation. Management should identify a strategy where knowledge can be transferred with speed, reliability and accuracy required in order to allow employees to get access to all kind of knowledge, regardless of whether the knowledge is available in or outside the organization. The results also shown as in Fig. 2 that organization culture, organization structure, technology and people have been found to be the highly significant KM enablers towards KM performance.

The hypothesis testing shows the satisfying estimate value for relationship between KM performance and organizational culture. It indicates that organizational culture has a positive relationship with KM performance in Malaysian public sector. In this case, particularly the individualism dimension as only items that reflect individualism were retained at the end of the analysis. This shows that words, information and knowledge are hardly ever shared across different units and different organisational levels. However, people share knowledge for some reasons such as reciprocity, reputation and prestige, do or sometimes just for altruistic reasons [13]. Therefore, management should develop a strategy that can change the attitude and behaviour of the people and reduce barriers because there can be no KM without knowledge sharing.

The finding for second hypothesis suggests that document confidentiality status and communication flow as the organizational culture dimensions are essential in influencing KM performance in organizations. For this study, organizational structure are only measured based on its influence on document confidentiality status and information flow. However, in terms of types of structure, the NRE has typical bureaucratic organizational form [26] which entails various vertical levels in organization. The result shows the estimate value of 0.296 for the relationship between KM performance and organizational culture. In the Ministry, documents and information are classified into four classifications, which are “open”, “confidential”, “secret” and “top secret”. The status of the documentation in the Ministry has an implication for the sharing of knowledge between individuals, divisions and organizations. However, Monavvarian and Kasaei [18] argue that if certain items of information and documents are restricted to certain levels of employee, it prevents the flow of knowledge across the organization. In addition, if an organization supports communication networks that operate freely, where knowledge providers and knowledge seekers can access information and knowledge through the shortest path, it will certainly enhance knowledge creation and knowledge transfer in the organization [5]. Therefore, the management should realize that it is very important for the knowledge to be available to all employees without any restriction, except for top secret documents. Management should always consider of improving this issue, especially in ensuring that the documents available in an organization can be accessed and shared.

The third hypothesis is to determine the significance between KM performance and technology. Both dimensions, ICT infrastructure and ICT tools have high influence towards KM performance. These shows that the officers beliefs that technology is one of the most important aspects that facilitate effective KM implementation. This tally with findings by Syed-Ikhsan and Rowland [5].

The fourth hypothesis is to determine the significance between KM performance and people. However, the results indicate that not all variables identified in the study have a relationship with the KM performance. Although the placements in the Ministry were suitable with the staff’s experience, interests and qualifications, the result shows no significant relationship with staff training and staff turnover. This might be due to the fact that the officers are yet to appreciate KM importance for their work.

Based on the outcome of the study, it is important for the top management of the organizations, particularly the NRE to have a clear, well-planned KM strategy before implementation can take its course. Although KM performance measurement could be different across organizational context [27], all the significant variables that have been tested and highlighted in this study should be taken into consideration. This will hopefully lead to a more effective KM practice for the organization.

References

1. Kuah, C.T., Wong, K.Y.: Knowledge management performance measurement : a review. Afr. J. Bus. Manag. **5**(15), 6021–6027 (2011). doi:10.5897/AJBM10.1091
2. Natarajan, G., Shekhar, S.: Knowledge Management: Enabling Business Growth. McGraw Hill, Malaysia (2001)
3. Handzic, M., Lagumdzija, A., Celjo, A.: Auditing knowledge management practices: model and application. Knowl. Manag. Res. Pract. **6**(1), 90–99 (2008). http://www.palgrave-journals.com/kmrp/journal/v6/n1/full/8500163a.html
4. Andone, I.I.: Measuring the performance of corporate knowledge management system. Informatica Economica **13**(4), 24–31 (2009)
5. Syed-Ikhsan, S.O.S., Rowland, F.: Knowledge management in a public organization : a study on the relationship between organizational elements and the performance of knowledge transfer. J. Knowl. Manag. **8**(2), 95 (2004)
6. Willem, M.A., Buelens, M.: Knowledge sharing in public sector organizations: the effect of organizational characteristics on interdepartmental knowledge sharing. J. Public Adm. Res. Theory **17**(4), 581–606 (2006). doi:10.1093/jopart/mul021
7. MAMPU: Knowledge Management Blueprint. MAMPU, Putrajaya (2011)
8. Chong, S.C., Salleh, K., Ahmad, S.N.S., Sharifuddin, S.-I.S.O.: KM implementation in a public sector accounting organization: an empirical investigation. J. Knowl. Manag. **15**(3), 497–512 (2011). doi:10.1108/13673271111137457
9. Lee, H., Choi, B.: Knowledge management enablers, processes and organizational performance : an integration and empirical examination. In: APDSI 2000
10. Bloodgood, J.M., Salisbury, W.D.: Understanding the influence of organizational change strategies on information technology and knowledge management strategies. Decis. Support Syst. **31**(1), 55–69 (2001)
11. Davenport, T.H., Prusak, L.: Working Knowledge: How Organizations Manage What They Know. Harvard Business Press, Cambridge (1998)
12. Cong, X., Pandya, K.V.: Issues of knowledge management in the public sector. J. Knowl. Manag. **1**(2), 25–33 (2003)
13. Yusoff, M.Y.M., Mahmood, A.K., Jaafar, J., Salak, K.K.P.: A study of KM process and KM enabler in a Malaysian Community College. J. Knowl. Manag. Pract. **13**(1), 27–33 (2012)
14. Erwee, R., Skadiang, B., Roxas, B.: Knowledge management culture, strategy and process in Malaysian firms. Knowl. Manag. Res. Pract. **10**(1), 89–98 (2012). doi:10.1057/kmrp.2011.37
15. Wei, C.C., Choy, C.S., Yew, W.K.: Is the Malaysian telecommunication industry ready for knowledge management implementation? J. Knowl. Manag. **13**(1), 69–87 (2009). doi:10.1108/13673270910931170
16. Nonaka, I., Toyama, R., Konno, N.: SECI, Ba, and leadership: a unified model of dynamic knowledge creation. Long Range Plan. **33**, 5–34 (2000)
17. Skadiang, B.: Dimensions of organisational knowledge management (OKM) a study on Malaysian Managers using the multidimensional USQ KM scale. University of Southern Queensland (2009)
18. Monavvarian, A., Kasaei, M.: A KM model for public administration: the case of labour ministry. Vine **37**(3), 348–367 (2007). doi:10.1108/03055720710825654
19. Rai, R.K.: Knowledge management and organizational culture: a theoretical integrative framework. J. Knowl. Manag. **15**(5), 779–801 (2011). doi:10.1108/13673271111174320
20. Bollinger, A., Smith, R.: Managing organizational knowledge as a strategic asset. J. Knowl. Manag. **5**(1), 8–18 (2001)

21. Bhatt, G.D.: Knowledge management in organizations: examining the interaction between technologies, techniques, and people. J. Knowl. Manag. **5**(1), 68–75 (2001)
22. Soon, T.T., Nyu, M.K.C., Jon, T.J., Fisher, J., Bakar, N.A., Tayib, M., Kwong, G.S.: Asian J. Knowl. Manag. Editorial Board. Asian J. Knowl. Manag. **1**(1) (2014)
23. Narayanan, Ramanathan, Richardson, S., Salleh, Abdul Latif: knowledge management: a malaysian study. Malays. Manag. Rev. **38**(1), 79–89 (2003)
24. Chen, C.J., Huang, J.W.: Strategic human resource practices and innovation performance—the mediating role of knowledge management capacity. J. Bus. Res. **62**(1), 104–114 (2009)
25. Awang, Z.: Structural Equation Modeling Using AMOS Graphic, 2nd edn. UiTM Press, Subang Jaya (2013)
26. Weber, M. The theory of social and economic organization. Simon & Schuster, New York (first published in German) (1922; 1997)
27. Wong, K.Y., Tan, L.P., Lee, C.S., Wong, W.P.: Knowledge management performance measurement: measures, approaches, trends and future directions. Inf. Dev. (2013). doi:10.1177/0266666913513278

A Study on the Influence of National Culture on Knowledge Sharing

Jouni A. Laitinen[1(✉)], Jan M. Pawlowski[2], and Dai Senoo[1]

[1] Industrial Engineering and Management, Tokyo Institute of Technology, Tokyo, Japan
{laitinen.j.aa, senoo.d.aa}@m.titech.ac.jp
[2] Hochschule Ruhr West University of Applied Sciences, Mülheim, Germany
jan.pawlowski@hs-ruhrwest.de

Abstract. The influence of national culture on knowledge sharing has important implications for all organizations. However, the existing frameworks only cover a subset of relevant factors or limit the research of the framework to either organizational or national level. Hence, a more encompassing framework is needed. The question this articles answers is how does national culture influence knowledge sharing. Based on extensive literature review and interviews carried out in Finland and Japan, this article sets forth a foundation for a new framework. The framework details how national culture influences individual level and organizational level factors and technical tools. Additionally, the framework includes a new dimension, time-dimension, which is usually disregarded in knowledge sharing research. For researchers and practitioners, the derived framework provides key insight on relevant factors on knowledge sharing and national culture. Finally, future research directions are discussed.

Keywords: Knowledge sharing · National culture · Intercultural sharing · Framework

1 Introduction

In this article we discuss the challenges organizations face to become "knowledge-generating, knowledge-integrating, and knowledge-protecting organizations" [42]. Knowledge sharing is one of the key activities within knowledge management [4]. During the last few years more and more researchers have started to acknowledge the influence of culture on knowledge sharing (c.f. [49]). However, usually researchers limit the scope of the research to a small subset of the possible aspects where the influence of national culture on knowledge sharing can be seen. Hence, the research question the researchers aim to answer is how does national culture influence knowledge sharing. This is done through an extensive literature review to understand relevant factors of the sharing process. We compare these factors to the results of interviews carried out in Finland and Japan. Based on the comparison of the literature review and the interviews, a framework is proposed for detailing the influence of national culture on knowledge sharing. Finally, the implications and limitations of the framework are discussed.

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 160–175, 2015.
DOI: 10.1007/978-3-319-21009-4_13

2 National Culture and Cultural Models

We define culture as behavioral patterns and traits, which are passed through social interaction and develop over time. In the field of knowledge management, culture can refer to two aspects: organizational culture and national culture [10]. In this article we concentrate solely on national culture while clearly acknowledging that both perspectives on culture are relevant. Mabawonku [27] defined national culture as "[…] definitive, dynamic purposes and tools (values, ethics, rules, knowledge systems) that are developed to attain group goals". However, previous research has found that not only is there significant variance in culture within a single country, there also exist significant overlap in culture between different countries [16]. In addition, using single site studies to study national culture implies a presupposition for a unified national culture [31]. Recently results, however, show that individual value dimensions and national value dimensions do not need to be studied separately [9]. This suggests that individual values and national values can be studied together and as such cultural models can be used to study national culture.

Cultural models can be used to gain a deeper understand of the dimensions of culture and what needs to be taken into account when studying the influence of national culture. While it can be argued that cultural models are simplifications of real culture [31], the models can still help to understand how the national culture differs between countries differs and what aspects are similar [9]. Within the field of knowledge management Hofstede's [16] Cultural Dimension Theory is the most widely used (e.g. [5, 28, 34]). In the original work Hofstede's dimensions consisted of power distance, uncertainty avoidance, individualism and masculinity. Later he added long-term orientation and indulgence-restraint dimension [18]. Hofstede's original work was carried out in IBM's offices in different countries. Hence, criticism has been presented about how the dimensions were derived and how applicable the results are ([31, 50]). However, recent work (e.g. [33]) still supports the cultural dimension model. Usually a subset of these values is used in knowledge sharing studies (e.g. [1, 5]). Another widely used cultural model is by Trompenaars and Hampden-Turner [45], whose model consists of seven dimensions. While their model covers some aspects of Hofstede's Cultural Dimension Theory, future research on the applicability of the extra dimensions in knowledge sharing domain should be investigated.

3 Models and Frameworks

The frameworks discussed in this section analyze relationships between national culture and various relevant aspects of knowledge sharing and knowledge transfer. The level of inspection of the frameworks varies between the frameworks and therefore it is up to the researcher to choose the applicable framework for that particular context. Bhagat et al. [3] use cultural patterns when analyzing effective knowledge transfer. Their proposed framework consists of two cultural dimensions: individualism-collectivism and verticalness-horizontalness. Bhagat et al. [3] also present a conceptual model detailing cross-border transfer of knowledge. The notable aspect of this framework is that national culture influences different types of knowledge and that individual factors also influence the success of transfer.

Huang et al. [20] theorized about knowledge sharing in China. The authors have taken a different approach to the previously presented framework, as their framework addresses only a single country instead of building a more abstract model. They considered face-saving, face-giving and guanxi orientation (social relationship orientation) as cultural factors. The results showed that the studied cultural factors influence knowledge-sharing intention significantly. One on hand, Huang et al. [20] researched typically Asian cultural features limiting the applicability of their framework. On the other, Huang et al. [20] were able to concentrate on more specific factors related to knowledge sharing in China making the model more applicable in the Chinese context.

Griffith et al. [15] proposed a conceptual framework detailing the relationship between relationship resource and knowledge resources, and how national culture influences the two. Their results indicate that national culture influences trust and commitment. Their results also indicate that national culture influences the links between trust, commitment and information sharing. However, the link between information resources and problem resolution was not influenced by culture. Given the other existing studies, this finding is quite unexpected. The researchers explain this unexpected finding through the existence of previously established norms.

As can be seen from the presented frameworks national culture seems to have an affect on knowledge sharing processes. The presented frameworks on how national culture influences knowledge sharing all take different approaches. Bhagat et at. [3] framework helps to understand how national culture influences knowledge types. Griffith et al. [14] gives perspective on how individual aspects and knowledge sharing are influence by culture. Finally, Huang et al. [20] create a very detailed framework for China. However, the presented frameworks do not offer a comprehensive description on the influence of national culture on knowledge sharing, as organizational, individual and knowledge level aspects are influenced by national culture. There is a need for a more encompassing framework.

4 Organizational Aspects

To better understand the how national culture influences knowledge sharing, an extensive literature review was carried out. The results have been divided into two main sections: organizational aspects and individual aspects. Organizational aspects influenced by national culture are discussed first.

Strach and Everett [40] stated that Western organizations are more likely than their eastern counterparts to have a **formal knowledge management strategy**. Additionally, they described that employees' job description are less defined in Japanese organizations than in western organizations. This vagueness in job description can be seen as a way to transfer tacit knowledge to between employees ([36, 40]). Identification with the company helps employees to create a common identity, which can help lessen barriers for sharing. Efforts to overcome sharing barriers would need to be customized since knowledge management strategies need to be customized to fit national culture [43].

Knowledge sharing initiatives need **managerial level support** for them to succeed ([30, 48]). By demonstrating the value of knowledge sharing initiatives in order to gain a competitive edge, employees are more willing to share. For managers to be able to encourage sharing, **trust** between members of the organization needs to be created and supported ([1, 30, 46]). This is particularly important for encouraging sharing of critical knowledge, as losing ownership of such knowledge can lead to fears of losing one´s employment ([35, 38]). National culture effects how trust through various ways. For example, the time needed to build trust is longer in collective cultures [17]. Increased trust makes it easier to overcome the fear of losing face [46].

Cultural competences of managers. Knowledge transfer between employees who share similar backgrounds and status were found to improve knowledge transfer [34]. Better understanding the influence of national culture to communication styles helps to improve sharing between employees of different background [34]. It is, therefore, important for managers to understand how to better facilitate communication between individuals on all levels of the organization [2]. Supporting communication lowers the barriers related to language and cultural problems. One approach to solve cultural and communication problems is through training programs ([7, 35]).

Incentives were found to increase knowledge sharing by Matsuo and Easterby-Smith [29] who studied Japanese companies. Similar results have been derived in the banking industry in Japan [24]. However, incentives have been found to have a negative effect overall on sharing [4] and the long-term effects of incentive use on knowledge sharing have been questioned ([25, 38]). Glisby and Holden [11] questioned the relationship between incentive use and willingness to share when comparing Western and Japanese employees, as in some situations Westerners seemed to need incentives while the Japanese did not need incentives. Nevertheless, it has been argued that organizations can have an effect on how willing individuals are to share creating correctly customized incentives [47].

Knowledge sharing needs access to **technical tools** and software, which support sharing between individuals and the organization. Willingness to use the technical tools is a key aspect but it needs support from the organization [13]. Organizations should encourage individuals to customize the tools used, as it improves knowledge management and sharing practices [6]. Customization of the tools used also increases tacit knowledge, which can be used to improve the other tools being used in the organization. Nonaka and Takeuchi [37] stated that currently used tools frequently display the influence of western culture. Formal and informal interaction support by the chosen tools is important for increasing sharing ([2, 13, 41]). Additionally, the fear of "losing face" due to language and other cultural problems can be lessened through tool customization [27].

As shown in this section, national culture influences numerous aspects of the organization. In order to understand how to facilitate differences, organizations need to understand how and where the influences can be seen and how to mitigate the effects.

5 Individual Aspects

Knowledge sharing cannot be forced [12], hence understanding how to facilitate and enable sharing is key for all organizations. This places the individual at the heart of knowledge sharing. It has been theorized that the national culture in **Japan emphases tacit** knowledge where as **western culture emphasis is on explicit knowledge** [11].

This difference impacts how individuals engage in sharing practices used to distribute knowledge. Westerners would place more emphasis on knowledge in explicit format i.e. documents where as Japanese would more likely share knowledge in unofficial meetings. Low individualism score of on the cultural dimension theory indicates that Japanese organizations and individuals are more likely to share knowledge within already established networks ([11, 28]).

Trust can help to overcome knowledge-sharing barriers ([1, 30]). Trust creation between individuals and organizations involved in the sharing process is essential for success. However, trust building in collective cultures requires more time [34]. As Usoro et al. [46] showed in their research, trust consists of multiple factors, i.e. competency-based, integrity-based and benevolence-based components, which need to be considered in relation knowledge sharing. Hence, trust needs to be considered also from the individual's perspective when implementing sharing initiatives.

I Increasing **Willingness to share** is important in order to increase actual sharing levels. Voelpel and Han [47] studied sharing in a foreign multinational corporation operating in China. Their results showed that the two main factors influencing the Chinese employees' willingness to share were incentives and national culture. It is therefore likely that proper customization of incentives by the organization can increase the amount of knowledge employees are willing to share. Willingness to share outside of the group is likely influenced by national culture since the norm of group orientation is learned through national culture. A shared cultural background helps to share knowledge due to having shared meaning [14].

An interesting aspect, which has not gained so much attention in the research community, is the different **perspective on improvement**. For example, Goh [13] theorized that Western cultures emphases on substantial change in contrast to Japan where emphasis is on constant improvement. This difference can be seen in for example the concept of kaizen in Japan, the idea of continuous improvement [21]. A related concept is the Wabi-Sabi, which encourages finding beauty in imperfection [44]. As trust lessens the effect of face [26], in a trusted organizational environment emphasizing kaizen and wabi-sabi aspects could encourage sharing of unfinished works and ideas.

In the last chapters it has been have shown that there are many other aspects of knowledge sharing, which are influence by national culture, but haven't been covered in any of the existing frameworks. The results show a clear gap in existing research. To better understand the influence of national culture on knowledge sharing, a framework covering the previously presented aspects should be created. Hence, the researcher took steps to fill in this gap in research through explorative interviews.

6 Interview Methodology

To better understand the relationship between knowledge sharing and national culture, the researchers chose to carry out interviews in Finland and in Japan. These two countries were selected because the national cultures of the two countries are considerably different, which makes understanding the difference in the influences easier. As existing research does not cover all relevant aspects of knowledge sharing, the researchers chose to approach the problem with an explorative approach. Hence, semi-structured interviews were selected as the research methodology. A semi-structured interview is a qualitative research method, where the order of the questions and the specific wording of the questions can be modified by the researchers based on the answers of the interviewee [39]. Interviewees were encouraged to express their views on the topics in addition to answering the questions asked by the researcher.

Candidate organizations both in Finland and in Japan were approached by email in order to encourage participation from multinational organizations operating in either or both of the countries. The aim was to recruit encourage both academics and industry individuals working in international settings to participate in the interviews. The industry individuals were employed in organizations ranging from import-export companies to innovation centers. The chosen academics were individuals with experience on topics related to knowledge sharing and national culture. The age range of the interviewees ranged from mid 20s to mid 60s. In total there were 9 interviews, which were divided into 4 Japanese and 5 Westerners. There were two academics from Finland and one from Japan. Industry interviewees consist of three interviews with Japanese individuals and three interviews with Finnish individuals.

The interviews were carried out both in Finland and in Japan during late April and early May of 2013. The interviews were taped and the researcher made notes during the discussion. The interviews with Japanese individuals were carried out in English and the interviews with Finnish individuals were carried out in Finnish. The length of the interviews ranged between 21 min and 62 min with an average length of 38 min. Two of the interviewees requested not to be taped. For those interviews extensive notes were taken during the discussion and unclear answers were clarified by further communication via email.

7 Results from the Interviews

Results from the interviews show support that national culture has an influence on knowledge sharing. The influence can be seen in multiple factors impacting the sharing process. In short, the influence of national culture can be seen in aspects of organizational and individual factors, trust, willingness to share and in the tools used.

Cultural aspects: The influence of **long-term orientation** is related to attitudes towards knowledge sharing, and particularly how organizations view time in proposed sharing practices. Based on the interviews, in Japanese organizations the time-scale of operations seems to be longer and knowledge-sharing initiatives are viewed on a longer time scale. This is clearly indicated by a quote from a Western individual (W1) working in industry in Japan, who quoted a Japanese proverb, which roughly translates as follows:

> “Peaches and chestnuts, three years; persimmon, eight years”.

In regards to sharing initiatives this means that Japanese organizations are more willing to let initiatives have time before concrete results are expected. Western organizations are less willing to wait for projects to bring for fruitful results and might stop sharing initiatives before they become successful. Long-term orientation is integrally related to trust building between members inside the organization as well as outside organizations.

The interviewees also discussed knowledge flow directions. The interviewees mentioned that the knowledge flows in Japanese organizations were more vertically oriented when compared to Finnish organizations. One Western individual (W2) working in an information technology organization located in Japan stated in relation to cultural influence to knowledge sharing in Japan:

"Correct knowledge comes from management".

This quotation clearly indicates that the top-down knowledge flow is automatically considered to be more "true" than communication flows, which have originated lower in the organizational structure. In comparison the interviewees working in Finnish organizations indicated that the organizational structures related to knowledge flows were more **horizontal** instead of the **vertical** structures of the Japanese organizations. These differences in knowledge flows seem to be related to the vertical and horizontal structure of the organizations and can pose barriers to sharing. Additionally, during the interviews numerous gave statements indicating that there was a more of an emphasis on **groups** in Japanese organizations than in Western organizations. This was clearly shown in a comment with a Finnish person, with a considerable background of working in Japan, who stated that sharing is done mostly within groups and previously established links in Japan. This in contrast to Finland where individuals were more willing to share also outside of the group.

Aspects related to individual: Aspects regarding language used to share knowledge that were mentioned by the interviewees include **similar level of language fluency** of the sharers. According to the interviewees, sometimes Japanese employees are less eager to share knowledge if that individual hasn't reached a certain level of fluency. When asked for a reason for this difference in willingness, a Japanese interviewee (J3) stated:

"Similarity in language skill level makes sharing easier".

Another reason stated by the interviewees was that lower eagerness to share might be due to wanting to avoid possible misunderstandings. Possible ways to help overcome this barrier is through having an **understanding of the cultural differences** between sharers as stated by multiple interviewees. Understanding of the differences in assumptions helps individuals communicate more clearly.

Trust was one of the key topics that interviewees brought up as a key factor influencing knowledge sharing and willingness to share with others. One Japanese industry individual (J2) condensed the role of trust in knowledge sharing into the following words:

"Without trust there is no knowledge sharing".

As members of the organization gain deeper levels of trust, they are willing to share more knowledge. This applies also to knowledge, which could give them an advantage if hoarded. However, if a certain level of trust has been established between members then knowledge sharing can take on the function of correcting false knowledge. As one Interviewee (W2) stated that

"Knowledge sharing is bullshit killing".

The elimination of false knowledge through knowledge sharing is most likely indicative that members are comfortable sharing between each other. However, it should be noted based on previously reported preferences on knowledge flow directions this use of knowledge sharing to rectify false knowledge might not be used as often in Japanese organizations if the false knowledge has originated from higher in the organizational hierarchy.

Organizational aspects: In this aspect the **language skills of the manager** are critically important so that she can delegate suitable tasks to her subordinates. In order to share knowledge on the tasks correctly, particularly for employee's who do not share the same background, the manager needs to have reached a certain level of fluency. Importance of language skills of the manager also applies when subordinates do not speak the common language as a native language. The manager needs to understand how to make herself understood by individuals who are less fluent in the common language. One way to achieving such a goal is to lower the level of the used grammar and vocabulary in such a way that the two parties use the same language level.

Regarding **incentive** usage to encourage sharing, many interviewees stated that there was no need for incentives in Japanese organizations. However, some Finnish interviewees stated a need for incentives to be used in Finnish organizations. This would indicate that there exists a difference in the need for incentive. It should be noted that two interviewees, one Western and one Japanese, stated that the need for incentives is more related to organizational culture than to national culture. In connection with the discussion on incentives, the targets of the incentives were also discussed. During the interviews more interviewees stated that the target of the incentive should be the group instead of individuals. A Japanese interviewee explained the rationality behind the group focus by stating that: "All individuals belong to a team and the team competes for incentives". The interviewees also brought up the importance of having a **common goal** to work towards, as this would increase willingness to share since in that scenario all members' efforts are directed at reaching one single goal instead of multiple individual goals.

During the interviews individuals working in western organizations were more likely to state that their organization had an explicit **knowledge sharing strategy**. As sharing policies require **management support** the same individuals were also more likely to state that management supported sharing. However, this does not mean that the managers in Japanese organizations were less supportive of sharing, as participation in the norm of sharing is expected in Japanese organizations. Therefore, there was less need for managers to explicitly support sharing. It should be noted that these two topics received less attention from the interviewees.

Practices the tools used for knowledge sharing also varied between the two countries. Finnish interviewees were more likely to state that their organization used wikis and other technical tools to help share knowledge throughout the organization. This was echoed by a Finnish interviewee working in a multination organization where email usage has started to take on more of an instant messaging style function. It is likely that the email system is seen as a convenient way to share knowledge between members. Japanese interviewees stated that email and meetings were the main way of sharing between departments. The knowledge flow directions could influence sharing practices some tools, such as wikis, can make it harder to distinguish the origin of the knowledge shared.

Finally, **willingness to share**. Many aspects of the previously discussed topics influence how willing individuals are to share knowledge. As previously quoted based on the interviews, trust is one of the key aspects in increasing willingness to share. Increasing the level of trust between the actors creates an environment, where sharing can be done without the fear of being judge or exploited. This is indicated by a quotation from a Finnish interviewee (W5), who stated that:

> "Trust increases willingness to share even critical information".

Second aspect influencing willingness to share is incentives. Based on the interviews it seems that the types of incentives used need to be carefully considered according to influence of the organization as well as the national culture. While the results on the need for incentives were mixed, the individuals who stated a need for incentives indicated that national culture could influence the incentives used. Third, **the sharing outside of the group**. Blocks in the knowledge flows are integrally related to sharing outside of groups, which gained also attention from the interviewees. It was noted by most of the interviewees that organizational policies limit sharing outside between organizations. However, there were also blocks on sharing within the organizations. Two of the interviewees stated that the existence of knowledge silos is mostly likely due to differences in knowledge flows. One of the Finnish interviewees working in Japan stated that knowledge silos existed particularly within organizations with ties to the Japanese governmental branches.

8 Framework for Detailing Cultural Influence Factors on Knowledge Sharing

Based on the interviews, there are factors, which are directly influenced by national culture, and factors, which are influenced indirectly by national culture. Therefore, a new framework detailing the influences and their relationships is proposed. The proposed framework consists of national culture, individual, organization, trust, willingness tools and knowledge sharing. Each of the factors consists of smaller attributes, which describe the influences in more detail. The framework and the relationship between the factors are shown in Fig. 1 Proposed framework.

While most of the factors and attributes are the same as the ones found during the literary review, a new connection and an attribute were also found. A new connection between national culture and willingness was found. The importance of a common goal set by the organization was also discovered based on the interviews. The attributes are shown in Table 1: Factors of the framework.

Since the scope of the proposed framework is wider than the any of the previously described frameworks (e.g. [32, 34]), the proposed framework can help researchers understand the influence of culture in a more complete way. This is due to the framework encompassing both individual and organizational level factors as well as how the factors link to each other.

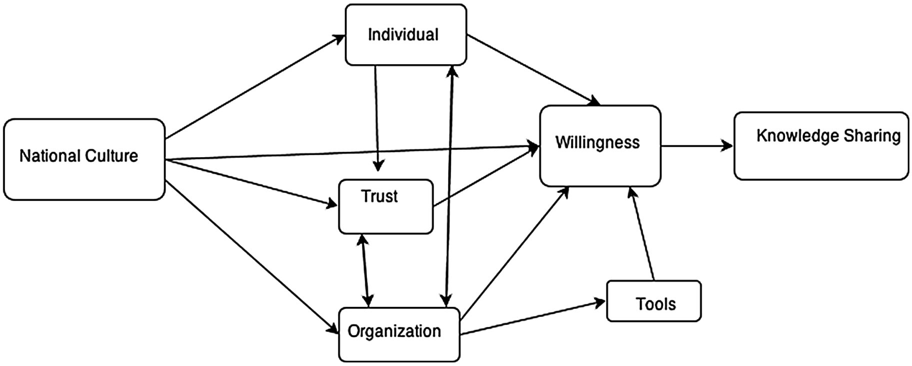


Fig. 1. Proposed framework

Table 1. Factors of the framework

Influencing factor	Influencing attributes	Links to	Supporting statements from interviewees
National culture	Individualistic -collectivistic Horizontal - vertical Long-term orientation	Individual Organization Trust Willingness	W1, W2, W3, W4, W5, J1, J2, J3, J4
Individual	Cultural distance Language fluency of the individual	Organization Trust Willingness	W4, W3, J2, J4
Organization	Incentives Management support for knowledge sharing practices Existence of formal knowledge sharing strategies Common goal	Individual Trust Tools Willingness	W1, W2, W3, W4, W5, J1, J2, J3, J4
Trust	Competency-based Integrity-based Benevolence-based	Willingness Organization Individual	W2, W4, W5, J2, J3
Tools	Support communication with other employees Ease of use	Willingness	W2, J3
Willingness	Sharing outside of group	Knowledge sharing	W3, W4, J3, J4

Cultural factors consist of **individualistic-collectivistic dimension**, **horizontal-vertical dimension** and **long-term orientation**. Having these three factors as the main sources of influence of national culture hasn't been previously proposed in knowledge sharing literature. First, **individualistic-collectivistic dimension** details the group-orientation of the employees. More group-oriented individuals are more likely to feel more open to sharing within the group ([34]). The interviewees also confirmed this. Second, **vertical-horizontal dimension** relates to the vertical structures in the organization. The more vertical the organization is the more likely the direction of knowledge flows to matter. For example, in Japanese organizations early feedback, and approval, is gotten outside of official meetings. Official meetings are used to approve initiatives and decisions that have been unofficially approved already. Third, **long-term orientation**. This final dimension is important for knowledge sharing as it is closely related to attributes of trust and how long sharing initiatives have until the organization considers cancelling the initiative. This dimension can be seen in the previously presented quote from an interviewee about peaches and chestnuts. Based on the interviews, these three cultural factors appear to be the main aspects of culture that influences knowledge sharing.

Individual factors consist of **cultural distance** and **language fluency of the individual**. **Language skills** and specially difference in the fluency are important influencing attributes in sharing knowledge. This derived attribute is inline with previous research [10, 35]. The interviewees indicated that by understanding how the change the level of the language used, individuals could increase sharing with others, who are less fluent communicators. **Differences in culture** can hinder communication flows [3]. The interviews supported this. Cultural distance can be another factor, as it is easier for individuals with similar backgrounds to share with each other [8, 14]. This was reported also by Cordeiro-Nilsson and Hawamdeh [7], who studied a Swedish company in Singapore and reported about the communication problems related to culture. The previously mentioned researchers suggested that cultural training could be done to help overcome this barrier for sharing. The interview the results regarding preferences between implicit and explicit were inconclusive.

Organizational factors include **incentives**, **management support for knowledge sharing practices**, **existence of formal knowledge sharing strategies** and **common goals**. One of the most controversial topics during the interviews was **incentiv**e usage. Both Japanese and Finnish interviewees stated opinions for and against the usage of incentives. Overall, it can be stated that Finnish interviewees were more likely to state a need for incentives than the Japanese interviewees. This result is consistent with previously derived results [23]. As stated by a Japanese individual working for a western organization, it seems that organizational culture is closely related to the need for incentives to encourage sharing, however the researchers were not able to confirm this during the interviews. The interviewees also discussed the targets of the incentives. One Japanese industry individual stated that they used groups as targets of incentives while western individuals reported the target to be individuals. This is inline with previously derived results [34]. Second big topic for the interviews was **common goals**. One Finnish industry individual stated that having a common goal is more important than any incentive system the organization could offer. Japanese interviewees echoed this. Having a common goal with management support can have a positive effect on sharing [41]. The last two factors are relatively intertwined as **sharing policies** require **management support** to work and management should support sharing between employees since increased collaboration between workers is usually beneficial. Based on the interviews, Japanese interviewees were less likely to state their organization had an official sharing policy. This is inline with Strach and Everett [40]. According to them the reason for this difference is the embeddedness of implicit knowledge sharing policies in the Japanese organization. This is shown in the

Japanese practices of job rotation as a way to share knowledge and to train employees. Leading by example is most likely the easiest way for managers to show support for both sharing policies and for sharing activities.

Trust factors are composed of three subcategories: **integrity-based trust**, **competency-based trust** and **benevolence-based trust**. As trust is influenced by both individual factors and organizational factors in addition national culture, it is shown as a separate entity in the framework. Based on the interviews, integrity-based and competency-based trust is more important than benevolence-based trust for knowledge sharing. **Integrity-based trust** and **competency-based trust** increase the amount of correct knowledge shared and **benevolence-based** trust ensures that the receivers of the knowledge do not abuse give knowledge. This can help over come some of the barrier for face in Asian cultures [43]. Overall it can be stated that the results are in line with Usoro et al. [46]. The influence of national culture on trust is most easily seen in how long it takes to build trust between members of the organization [16]. In Japanese organizations the new employee becomes a member of the organizational family, which is in contrast to Finnish organizations where new employees are hired mostly based on competence. Hence, the first type of trust used is different between the two countries. Organizations should support trust-building through positive experiences as a way to increase sharing ([26, 29, 35]).

Out of all the topics discussed during the interviews, **tools** gathered the least amount of attention. Based on the interviews the two most important aspects regarding knowledge sharing are **ease of use** and **support for communication**. The **ease of use** is important since if sharing via tools takes more effort than employees are less likely to use the tools for sharing [22]. **Support for communication** There were indications of differences in usage patterns of electronic tools in the interviews. For example, Western interviewees were more likely say that their organization used a wiki. Previous research has also shown that cultural influences have an impact on email usage patterns [19]. The interview results on the preference between substantial and incremental change were inconclusive. Future research examining knowledge repository usage could result in more conclusive findings on this aspect.

Willingness to share takes influence from multiple previously discussed factors in addition to having an attribute within it. Most of the influences for willingness to share come from national culture, individual, organization and trust. The only attribute within willingness to share is **willingness to share outside of group**. As there are multiple influencing factors, trust as a separate entity. Willingness to share outside of group is influenced by national culture's individualism-collectivistic dimension. Number of the Finnish interviewees mentioned that they had encountered more knowledge silos within Japanese organizations that in Western organizations. Willingness to share outside of group was also mentioned when discussing inter-organizational sharing. The results in regards to willingness to share outside of the group are inline with previous research [3]. Trust is likely to be the most important factor influencing willingness to share outside of group. Usoro et al. [46] stated that trust increase sharing of even critical knowledge. The interviewees also corroborated this. Hence, the link between trust and willingness to share is most likely the most important influence relationship out of all of the influencing factors to willingness. Individual factors such as cultural distance between the sharer and the receiver, and language skills can influence willingness to share since if the relationship is strained by communication difficulties, it is likely that increasing the amount of knowledge shared is difficult. However, if the two parties share similar level of a common background or a common language then sharing is easier and thus more frequent. Organizational influences can come from organizational practices such as incentives and establishment of common goals. By creating properly customized incentives to suit the sharer's preferences, it is likely that willingness to share is also increased. Common goals can also help to increase willingness to share, as then the employee is not only evaluated based on her work but based on the group output.

Based on the interviews, an original framework has been proposed to detail the ways national culture influences knowledge sharing. The proposed framework covers both individual and organizational level factors. By understanding the framework, researchers can gain a more thorough understanding of how to correctly customize sharing initiatives in the target organizations and country. Improving the antecedes of willingness to share is the easiest way to encourage sharing.

9 Conclusions, Limitations and Future Work

In this article we have discussed the link between national culture and knowledge sharing. First national culture and cultural models were discussed. Then a literary review of existing research on the connection between national culture and knowledge sharing was presented. The review showed a clear gap in existing frameworks. To fill the gap, interviews were carried out to understand how practitioners and academics understand the influence of national culture on knowledge sharing. Finally, based on the interviews, a new framework detailing the influence of national culture on knowledge sharing was proposed. The proposed framework depicts a more encompassing approach to knowledge sharing and national culture as it covers factors from the individual level as well as the organizational level. The proposed framework adds to research on knowledge sharing and national culture in the following ways: first, based on the interviews it can be stated that more cultural factors need to be considered. Increasing the number of cultural dimensions to three helps the research community understand that in addition to horizontal-vertical structures and individualism-collectivism –dimension time scale also needs to be considered. To the researchers knowledge no previous framework on knowledge sharing and national culture has included the combination of these three dimensions. Second, the proposed framework is not limited to depicting only individual factors or organizational factors. Instead the proposed framework depicts both of these influences. The combination of the first two contributions enables better customization of knowledge sharing initiatives. Third, the inclusion of technical tools to the framework. While the tools itself is not directly influenced by national culture, the way the tools are designed to be used needs to consider differences in communication styles and preferences in usage. These three contributions represent new directions for possible future research in aspects of initiative creation and tool customization.

The proposed framework has some limitations that need to be taken into account when applying it. First, the framework was derived through explorative interviews and literature review. However, due to the total number of individuals interviewed there could exist other aspects of national cultural influences on knowledge sharing which were left undiscovered. The limited number of interviewees can also affect the generalizability of the derived framework. In order to validate the model there is a need to carry out more interviews and a quantitative study testing the framework.

Future work based on the framework should concentrate on verifying the model based on quantitative methodologies. The research should be done as comparative studies between countries. This would validate the model for further use by other researchers and practitioners. Additional subsections of the model are fruitful for further research. For example, factors influencing technical tool customization, the preference between explicit and tacit knowledge and perspectives on improvement could all be targeted with a focus study based on a deeper analysis of knowledge repository usage in multinational organizations.

References

1. Al-Alawi, A.I., Al-Marzooqi, N.Y., Mohammed, Y.F.: Organizational culture and knowledge sharing: critical success factors. J. Knowl. Manag. **11**(2), 22–42 (2007)
2. Alin, P., Taylor, J.E., Smeds, R.: Knowledge transformation in project networks: a speech act level cross-boundary analysis. Proj. Manag. J. **42**(4), 58–75 (2011)
3. Bhagat, R.S., Kedia, B.L., Harveston, P.D., Triandis, H.C.: Cultural variations on the cross-border transfer of organizational knowledge: an integrative framework. Acad. Manag. Rev. **27**(2), 204–221 (2002)
4. Bock, G.-W., Kim, Y.-G.: Breaking the myths of rewards: an exploratory study of attitudes about knowledge sharing. Inf. Resour. Manag. J. (IRMJ) **15**(2), 14–21 (2002)
5. Boh, W.F., Nguyen, T.T., Xu, Y.: Knowledge transfer across dissimilar cultures. J. Knowl. Manag. **17**(1), 29–46 (2013)
6. Cabitza, F., Colombo, G., Simone, C.: Leveraging underspecification in knowledge artifacts to foster collaborative activities in professional communities. Int. J. Hum. Comput. Stud. **71** (1), 24–45 (2013)
7. Cordeiro-Nielsson, C.M., Hawamdeh, S.: Leveraging socio-culturally situated tacit knowledge. J. Knowl. Manag. **15**(1), 88–103 (2011)
8. Darr, E.D., Kurtzberg, T.R.: An investigation of partner similarity dimensions on knowledge transfer. Organ. Behav. Hum. Decis. Process. **82**(1), 28–44 (2000)
9. Fischer, R., Poortinga, Y.H.: Are cultural values the same as the values of individuals? An examination of similarities in personal, social and cultural value structures. Int. J. Cross Cult. Manag. **12**(2), 157–170 (2012)
10. Ford, D.P., Chan, Y.E.: Knowledge sharing in a multi-cultural setting: a case study. Knowl. Manag. Res. Pract. **1**(1), 11–27 (2003)
11. Glisby, M., Holden, N.: Contextual constraints in knowledge management theory: the cultural. Embeddedness of Nonaka´s knowledge-creating company. Knowl. Process Manag. **10**(1), 29–36 (2003)
12. Gibbert, M., Krause, H.: Practice exchange in a best practice marketplace. In: Davenport, T.H., Probst, G.J.B. (eds.) Knowledge Management Case Book: Siemens Best Practices, pp. 89–105. Publicis Corporate Publishing, Germany (2002)

13. Goh, S.W.: Managing effective knowledge transfer: an integrative framework and some practical implications. J. Knowl. Manag. **6**(1), 23–30 (2002)
14. Grant, R.M.: Toward a knowledge-based theory of the firm. Strateg. Manag. J. **17**(S2), 109–122 (1996)
15. Griffith, D.A., Myers, M.B., Michael, G.: An investigation of national culture ' s influence on relationship and knowledge resources in interorganizational relationships between Japan and the United States. J. Int. Mark. **14**(3), 1–32 (2006)
16. Hofstede, G.: Cultures Consequences: International Differences in Work-Related Values. SAGE Publications, Beverly Hills (1980)
17. Hofstede, G.: Cultures and Organizations: Software of the Mind. McGraw-Hill, London (1991)
18. Hofstede, G., Hofstede, G.J., Minkov, M.: Cultures and Organization (Rev. 3rd ed.). McGraw-Hill, New York (2010)
19. Holtbrugge, D., Weldon, A., Rogers, H.: Cultural determinants of email communication styles. Int. J. Cross Cult. Manag. **13**(1), 89–110 (2012)
20. Huang, Q., Davison, R.M., Gu, J.: Impact of personal and cultural factors on knowledge sharing in China. Asia Pac. J. Manag. **25**(3), 451–471 (2008)
21. Imai, M.: Kaizen: the Key to Japan´s Competitive Success. Random House Business Division, New York (1986)
22. Kankanhalli, A., Tan, B.C.Y., Wei, K.-K.: Contributing knowledge to electronic knowledge repositories: an empirical investigation. MIS Q. **29**(1), 113–143 (2005)
23. Kidd, J.: Working together, but how? The need for intercultural awareness. In: Beecher, S. L., Bird, A. (eds.) Japanese Multinationals Abroad: Individual and Organizational Learning, pp. 211–234. Oxford University Press, New York (1999)
24. Kubo, I.: An inquiry in the motivation of knowledge workers in the Japanese financial industry. J. Knowl. Manag. **6**(3), 262–271 (2002)
25. Lin, C.-P.: To share or not to share: modeling tacit knowledge sharing, its mediators and antecedents. J. Bus. Ethics **70**(4), 411–428 (2007)
26. Luna-Reyes, L.F., Cresswel, A.M., Richardson, G.P.: Knowledge and development of interpersonal trust: a dynamic model. In: Proceedings of the 37th Hawaii International Conference on System Sciences, pp. 5–8 (2004)
27. Mabawonku, A.O.: Cultural framework for the development of science and technology in Africa. Sci. Public Policy **30**(2), 117–125 (2003)
28. Magnier-Watanabe, R., Senoo, D.: Shaping knowledge management: organization and national culture. J. Knowl. Manag. **14**(2), 214–227 (2010)
29. Matsuo, M., Easterby-Smith, M.: Beyond the knowledge sharing dilemma: the role of customization. J. Knowl. Manag. **12**(4), 30–43 (2008)
30. McBeath, A., Ball, P.: Towards a framework for transferring technology knowledge between facilities. Strateg. Outsourcing Int. J. **5**(3), 213–231 (2012)
31. McSweeney, B.: Hofstede´s model of national cultural differences and their consequences: a triumph of faith – a failure of analysis. Hum. Relat. **55**(1), 89–118 (2002)
32. Michailova, S., Hutchings, K.: National cultural influences on knowledge sharing: a comparison of China and Russia. J. Manag. Stud. **43**(3), 383–405 (2006)
33. Minkov, M., Hofstede, G.: Hofstede's fifth dimension: new evidence from the world values survey. J. Cross Cult. Psychol. **43**(1), 3–14 (2010)
34. Möller, K., Svahn, S.: Crossing east-west boundaries: knowledge sharing in intercultural business networks. Ind. Mark. Manag. **33**(3), 219–228 (2004)
35. Noll, J., Beecham, S., Richardson, I.: Global software development and collaboration: barriers and solutions. ACM Inroads **1**(3), 66–78 (2010)
36. Nonaka, I.: The knowledge-creating company. Harvard Bus. Rev. **69**(6), 96–104 (1991)

37. Nonaka, I., Takeuchi, H.: The Knowledge-Creating Company. Oxford University Press, New York (1995)
38. Riege, A.: Three-dozen knowledge-sharing barriers managers must consider. J. Knowl. Manag. **9**(3), 18–35 (2005)
39. Robson, C.: Real World Research, 2nd edn. Blackwell Publishing, Malden (2002)
40. Strach, P., Everett, A.M.: Knowledge transfer within Japanese multinationals: building a theory. J. Knowl. Manag. **10**(1), 55–68 (2006)
41. Taminiau, Y., Smit, W., de Lange, A.: Innovation in management consulting firms through informal knowledge sharing. J. Knowl. Manag. **13**(1), 42–55 (2009)
42. Teece, D.J.: Strategies for managing knowledge assets: the role of firm structure and industrial context. Long Range Plan. **33**(1), 35–54 (2000)
43. Tong, J., Mitra, A.: Chinese cultural influences on knowledge management practice. J. Knowl. Manag. **13**(2), 49–62 (2009)
44. Treviranus, J.: The value of imperfection: the Wabi-Sabi principle in aesthetics and learning. In: Open ED 2010 Proceedings. OUC, OU, BYU, Barcelona (2010)
45. Trompenaars, F., Hampden-Turner, C.: Riding the Waves of Culture: Understanding Diversity in Global Business, 2nd edn. McGraw-Hill, New York (1998)
46. Usoro, A., Sharratt, M.W., Tsui, E., Shekhar, S.: Trust as an antecedent to knowledge sharing in virtual communities of practice. Knowl. Manag. Res. Pract. **5**, 199–212 (2007)
47. Voelpel, S.V., Han, Z.: Managing knowledge sharing in China: the case of Siemens ShareNet. J. Knowl. Manag. **9**(3), 51–63 (2005)
48. Zheng, W., Yang, B., McLean, G.N.: Linking organizational culture, structure, strategy, and organizational effectiveness: mediating role of knowledge management. J. Bus. Res. **63**(7), 763–771 (2010)
49. Wang, S., Noe, R.A.: Knowledge sharing: a review and directions for future research. Hum. Resour. Manag. Rev. **20**(2), 115–131 (2010)
50. Williamson, D.: Forward from a critique of Hofstede's model on national culture. Hum. Relat. **55**(11), 1373–1395 (2002)

The Relationship Between Knowledge Sharing Climate and Conflict Resolution Styles

Rasha Abou Samra(✉) and Khaled Shaalan

British University in Dubai, Dubai, UAE
rabousamra@hct.ac.ae, khaled.shaalan@buid.ac.ae

Abstract. This research is a case study research that is trying to solve the problem of task conflict that arises within project teams among university students. The research tests the relationship between knowledge sharing climate within the university and conflict resolution styles. This research used interviews with 10 students to explore the research questions and another interview with the head of the quality assurance department in the university that is responsible about the quality of the educational process in the university. Some indicators on the need for such research were collected through the documentation of the university records which was published on its website. A survey also of 102 students revealed that there is positive relationship between the knowledge sharing climate and the middle and high levels of collaboration and assertiveness in solving task conflicts that arise among students' team members. Avoidance which represents the lowest assertive and the lowest collaborative conflict solution is not related to the knowledge sharing climate at all in this research. This reflects that avoidance may lead to lack of learning unless it is related to dysfunctional conflicts. The research is an exploratory study for predicting the levels of collaboration and assertiveness in conflict resolution styles by the student's perception of knowledge sharing culture.

Keywords: Knowledge sharing climate · Conflict resolution styles

1 Introduction

This research is a try to fill the gap in the knowledge in the relationship between the conflict resolution styles and the knowledge sharing climate. The research is applied on university students who have projects to be implemented as a requirement to pass their courses. During the implementation of those projects students go through stages of group formulation and face some conflicts. They have different styles of solving those conflicts. Some collaborate with their colleagues and others avoid interaction with colleagues. Is there any significant relationship between the knowledge sharing climate in the university and the conflict resolution styles used by those students? This research is an attempt to explore the answer of this research question.

Project teams are formulated by passing through different stages: forming, Storming, norming, performing, and completion. During storming stage conflicts arise and the collaboration is in its minimal levels. The team listening is weak and some members of the team start avoiding challenges and return back to their comfort zones.

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 176–189, 2015.
DOI: 10.1007/978-3-319-21009-4_14

This is accompanied with distrust, conflicts, and lack of transparency. This lack of transparency in the researcher's point of view means lack of knowledge sharing. Students in universities have repetitive chances to experience storming in their project assignment teams. They also experience the high connectivity supported by technological social networks. They have an atmosphere of semi-equality in their demographic dimensions which may or may not enhance their knowledge sharing levels and their functional productive task conflicts to perform their project assignment perfectly. This research is a try to reach a statistically tested model to explain the relationship between the knowledge sharing climate within universities and the level of collaboration and assertiveness in resolving project assignments conflicts within a case study university.

2 Literature Review

The Knowledge collecting and the knowledge donating are interrelated terminologies to each other. (Hooff and Ridder 2004) found that the communication climate and the computer – mediated communication are positively influencing the knowledge sharing process. Another research done in (2007) by Lin found that the factor of enjoying helping others is significantly influencing the knowledge sharing process. The research found also that the person's willingness to collect and to share knowledge affects the improvement of innovation capability. The researcher found that it is very important to find out the knowledge sharing climate in the university context to support the innovative outcomes of learning and educational processes. Another research by (Gupta 2008) revealed that trust is one of the factors that encourage knowledge sharing. The research points at the knowledge sharing as a reward by itself, however, it did not point that there is a need to reward those who are more active in knowledge sharing. In the educational sector a research is needed to find out the knowledge sharing climate components among students to be able to reinforce it by university policies and learning tools and techniques. Conflict resolution was one of the main research topics within organizational contexts; however it did not take the same importance within students' universities contexts. There is a strong relationship found by researchers between knowledge sharing and collaboration (Rousseau 1998).

On the other hand researchers found that there is strong relationship between collaboration also and conflict resolution (Leung 2008). When there is a conflict between two parties, each one of them is thinking mainly about two things. The first thing is his/her own interests and the second thing is caring about the interest of the other party in the conflict. Accordingly, researchers found 5 styles of dealing with conflict based on those two dimensions. The first party may take care only with his/her own interests regardless of the other party's interest. This kind of people is highly assertive and is very weak in collaboration with the other party. Another style is the opposite to the first one, where we can find that one party is taking care of the other party's interest regardless of his/her own interest and this is representing high collaboration and low assertiveness style of resolving conflict situations. There are some people in the middle of taking care about their own interests and even about the other party's interest and they show middle levels of both collaboration and assertiveness.

Some people are low in both. Simply they do not care about anything, neither about their own interests, nor about other's interests in the conflict situation. They are careless and they do not think that conflict itself is acceptable so they avoid conflict situations. The last style of conflict resolution is the best one which is representing the person who is highly caring about his/her interest and at the same time highly caring about the other party's interest. As a conclusion researchers found 5 styles for resolving conflict situations. They are avoiding "low assertiveness and low collaboration", accommodating "high collaboration and low assertiveness", compromising "in the middle of both assertiveness and collaboration", competing "high assertiveness and low collaboration", and finally collaborating "high in both assertiveness and collaboration". (Thomas and Kilmann 1974).

The literature revealed that conflict is not always harmful and avoidable by people. Jehn found that conflict avoidance is ineffective as the same as what was found by Barker. Task conflict refers to the incompatible views, ideas, and opinions among group members about the content of their decisions (Han and Harms 2010). Leung (2008) found that conflict resolution cooperation strategy is a constructive problem solving conflict strategy and that it is able to build trust. Trust is one of the variables that are strongly and positively related to knowledge sharing as many researchers found (Lewicki et al. 1996). Task conflict rather than relationship conflict leads to more effective teams (Jehn 1997). This task conflict within universities and among university students is expected to include their disputes about their project assignments and how to perform the task. Task conflict is more cooperatively discussed by researchers and more contributing to the effectiveness of the team, however teamwork within organizational context may or may not differ from a team of learners within the university.

A research is needed to find how collaboration links between conflict and knowledge sharing. Knowledge sharing is easier when it is related to work tasks but how does it interact with possible task conflicts that may arise during work? More explicit knowledge is available in performing the task than in relational interactions. Boros et al. (2010) found that if people are strongly connected to each other and feel equality when they work together, they tend to use more collaborative conflict resolution strategies and less avoidance conflict resolution strategies and vice versa. If less connectivity and more inequality is experienced, then more avoidance conflict resolution strategies are used and less collaborative conflict resolution strategies are used. The same assumption need to be tested within university students. This high connectivity is expected to be found more within the students' environment. Open communication was found to be more innovative as well as the functional task conflict (Lu et al. 2011; Yesilbas and Lombard 2006).

3 Research Questions

Based on the literature review, the secondary data collected from the university published documentation, and the primary data collected through structured interviews, the researcher formulates the research major questions as follows:

"What is the significance of the relationship between the student's perception of the university knowledge sharing climate and his/her style of resolving studying projects' conflicts? And how are those variables related to each other? Does knowledge sharing climate contribute to the prediction and explanation of the variance in the task conflict resolution assertiveness and collaboration levels among university students?"

4 Research Objectives

1. To describe the relationship between conflict resolution styles and the general knowledge sharing climate within the University of Sharjah.
2. To describe the relationship between knowledge sharing components and conflict resolution styles used by university students in the case of this research.
3. To build a tested model of relationships between the case study variables that researchers can test on more case studies and to help decision makers in the University of Sharjah to improve the knowledge sharing climate which leads to improvement in educational outcomes' innovation.
4. To predict and explain the variance in conflict resolution style by the variance in the student's perception of the knowledge sharing climate within the university.

5 Research Methodology

This research is a case study that is conducted through semistructured interviews and surveys for collecting evidence information to reach the case study conclusion. The interview points are attached in the appendixes. Those primary data are also supported by secondary data coming from the website of the University of Sharjah. University of Sharjah programs are nationally and internationally accredited by different accreditation bodies like the Ministry of Higher Education and Scientific Research in the UAE, The engineering Accreditation Commission of ABET, the International Federation of Surgery and Obesity, Metabolic Disorders (IFSO), the European Association for Endoscopic Surgery (EAES), and the Commission for Academic Accreditation (CAA) in the United Arab Emirates. Sampling unit of this research is the University of Sharjah undergraduate students during the research data collection period.

A questionnaire was designed based on the literature review to measure the relationship between the knowledge sharing climate and the conflict resolution style. 5 point likert scale is used in the questionnaire. The items of the questionnaire were refined according to the judgment of three experts in the field. The questionnaire then was tested on a pilot sample of 10 respondents. Cronbach's Alpha for the pilot sample is 0.726. The researcher collected 102 questionnaires from undergraduate students via electronic surveys from University of Sharjah. The histogram showed the natural distribution of the sample responses. The value of Cronbach's Alpha for the whole sample is 0.791. All respondents both from interviews or questionnaires were Arabian students. The research questionnaire was designed in English, translated into Arabic, and after taking the approval of the same three experts from the field, it was back translated in to English language again. Sample of the questionnaire is attached in Appendixes.

The sample descriptive statistics shows that out of 102 students 49 were males, 51 were females. Regarding the academic level 22 were accepted, 39 were good, 30 were very good, and 11 were excellent. Regarding the number of years spent in this university 33 have been studying in this university for 1–2 years. 49 students have been studying in this university for 3–4 years, and 20 students have been studying in this university for more than 4 years. Regarding the age of respondents 25 students were aged 16 to less than 18, 43 were aged 18 to less than 20, 34 were aged 20 or more.

The knowledge sharing climate is measured by 4 factors in this research. Social gatherings, accommodation, encouragement of more clever students, culture of wishing success, rewarding those who share knowledge, willingness to share knowledge, closeness of students to each other, and the culture of accepting others' opinion without blaming, are all belonging to one factor which is factor number 1. The researcher calls this factor *the perception of knowledge sharing culture*. The second factor that measures the knowledge sharing climate in the college is consisting of how students perceive helping others to learn regardless of academic status, enjoyment of helping others, perception of learning from group discussions and from colleagues, willingness to share personal experiments, preferring to work with others rather than to work alone, freedom of accepting conflict in opinion, and openness of communication in general. The research calls this factor *the perception of communication openness*. The third factor in measuring the knowledge sharing climate is consisting of *the perception of the knowledge sharing conducive*. This factor consists of 2 conducive reasons; the first one is the perception of the importance of the knowledge the students would like to share and the second one is the perception of the level of trust with students we share the knowledge with. The last factor is measuring the hindrance items of knowledge sharing. They consist of the inability to increase connections by using the internet social networks, the inability to share project's knowledge via social networks, the inability to share knowledge with students with higher GPA, and the inability to share personal knowledge. The first and the second items are recoded to measure the inability. The third and the forth are left as they are to measure the inability. Those items are forming the fourth factor in the knowledge sharing climate. The researcher calls this factor the *hindrances of knowledge sharing*. The knowledge sharing climate in this research consists of the following 4 items: *1- The perception of knowledge sharing culture, 2- The perception of communication openness, 3- The knowledge sharing conducive, & 4- Hindrances of knowledge sharing*. Factor analysis showed 5 factors of conflict resolution styles in this sample. Factor 1 consists of 2 items and is measuring the compromise conflict resolution style. The second factor consists of 2 items and is measuring the accommodation conflict resolution style. The third factor also consists of 4 items and is measuring cooperative conflict resolution style. The fourth factor consists of 2 factors also, and is measuring the competitive conflict resolution style. The fifth factor consists of 2 items and is measuring the avoidance conflict resolution style (Table 1).

The one – sample Kolmogorov-Smirnov test results showed that the test distribution is normal as shown by the following Table 2.

Table 1. Factor analysis for conflict resolution styles – Equamax rotation with Kaiser Normalization.

Items	Factor 1 Compromise	Factor 2 Accommodation	Factor 3 Collaboration	Factor 4 Competition	Factor 5 Avoidance
During conflict, I try to reach a compromise to satisfy all parties	0.562				
When we have different points of view, I always try to find a common area	0.513				
During the conflict, I have to take the side of one party to win the conflict		0.697			
I avoid those colleagues who have strong opinion		0.542			
In conflict situations, I cooperate with my colleagues and accept their instructions easily			0.548		
I am always willing to listen to my colleagues' opinion, but I am also willing to give them my opinion			0.521		
When a conflict arises within my project group I am always willing to modify my priorities to solve the conflict			0.596		
I make my decisions by myself, but I listen carefully to other's opinion to reach the best decision			0.522		
I always consider my colleagues opinion, but I always make my decisions by myself				0.653	
When conflict arises in my project group, I always stick to my principles				0.703	
If my colleagues do not respect my opinion , I keep it to myself					0.605
When a conflict arises, I tend to avoid it and care about something else					0.532

Table 2. Test of normality for research factors

Factor	Z value	Sig. 2 tailed
Knowledge sharing culture	0.658	0.780
Communication openness	1.008	0.262
Knowledge sharing conducive	1.670	0.008
Hindrance of knowledge sharing	1.028	0.241
Compromise conflict resolution	1.867	0.002
Competition conflict resolution	1.852	0.002
Avoidance conflict resolution	1.602	0.012
Accommodation conflict resolution	1.374	0.046
Collaboration conflict resolution	1.007	0.263
General knowledge sharing climate	0.992	0.279

6 Hypotheses Testing

Based on the research objectives and the research questions the researcher will test the following hypotheses using the Table 3.

H1: There is a significant positive relationship between student's General perception of knowledge sharing climate and their styles of conflict resolution.

To test this hypothesis the researcher conducted pearson correlation test among the research factors of the knowledge sharing climate and the conflict resolution styles and it is statistically valid to say that the change in the perception of the knowledge sharing

Table 3. Correlations for Hypotheses one, two, and three.

Items	Conflict resolution styles	Correlation coefficient	Sig. (2 tailed)
General perception of knowledge sharing climate	Compromise conflict resolution style	0.351	0.000
	Competition conflict resolution style	0.376	0.000
	Accommodation conflict resolution style	0.224	0.024
	Collaboration conflict resolution style	0.249	0.012
Items	Conflict resolution styles	Correlation coefficient	Sig. (2 tailed)
Perception of knowledge sharing culture	Compromise conflict resolution style	0.203	0.041
	Accommodation conflict resolution style	0.332	0.001
	Competition conflict resolution style	0.247	0.01
	Conflict resolution styles	Correlation coefficient	Sig. (2 tailed)
Perception of communication openness	Compromise conflict resolution style	0.321	0.001
	Competition conflict resolution style	0.373	0.000
	Collaboration conflict resolution style	0.291	0.003

climate in the university is related significantly to the change in using conflict resolution styles by university students. It is also found that both dimensions are changing in the same direction. The positive perception of knowledge sharing climate is related to higher tendencies to use compromise, competition, accommodation, collaboration conflict resolution styles. There was no significant relationship between avoidance conflict resolution style and perception of knowledge sharing climate. This means that the change in using the avoidance style is not related to the change in the perception of knowledge sharing among university students. According to this interpretation of hypothesis testing results we partially accept this hypothesis and partially reject it only regarding the avoidance conflict resolution style.

H2: There is a significant relationship between the Arabian perception of knowledge sharing culture and the conflict resolution styles used by students belonging to this culture.

According to the table we can conclude that the change in the perception of the knowledge sharing culture and the change in using conflict resolution styles are positively and significantly related. Increased compromise, accommodation, and competition conflict resolution styles are related to the perception of more positive knowledge sharing culture and vice versa. More or less avoidance or more or less collaboration is not significantly related to the perception of knowledge sharing culture. This is only applicable on this case study and not generalized in all universities. If we would like to reach generalization then more case studies should be included to make representative sample of UAE universities or Arabian Universities. Again we can say that this hypothesis is partially accepted regarding the relationship between perception of Knowledge sharing culture and compromise, accommodation, and competition conflict resolution styles. We also conclude that this hypothesis is partially rejected regarding the avoidance and collaboration conflict resolution styles and their relationship with the perception of knowledge sharing culture for university Arab students.

H3: There is a significant positive relationship between the variance in the perception of communication openness in the university and the variance in using conflict resolution styles.

Based on the data in the table we found that the change in the perception of communication openness is significantly related to the change in using compromise, competition, and collaboration conflict resolution styles. Again this result is not generalized to all university students. It is only true regarding the sample units of this case study. Change in the avoidance and accommodation style of conflict resolution is not related to the change in perception of communication openness for this sample. We can conclude that this hypothesis is partially accepted regarding the relationship between perception of communication openness and compromise conflict resolution style, competition resolution style, and collaboration resolution style. The hypothesis is also partially rejected regarding the relationship between perception of communication openness and avoidance conflict resolution style and accommodation conflict resolution style.

H4: the student's perception of the knowledge sharing climate in the university has a significant positive effect on the variance in conflict resolution style of this student.

H4A: Student's perception of the level of communication openness as a component of knowledge sharing climate in the university affects significantly and positively his/her level of using cooperative task conflict resolution style.

To test this hypothesis the researcher used regression test and found the following.

Table 4. Regression model for hypothesis 4A

Variables entered	R	R square	Adjusted R square	Std. Error of the estimate	F	Sig.
Perception of communication openness	0.291	0.085	0.076	2.45000	9.281	0.003

Dependent variable is collaboration conflict resolution style.

– The model is significant and the coefficients are as follows.

Table 5. Coefficients of regression model for H4A

Model	Unstandardized coefficients		Standardized coefficients	t	Sig.
	B	Std. Error	Beta		
Constant	7.123	0.981		7.259	0.000
Perception of communication openness	0.140	0.046	0.291	3.046	0.003

The Tables 4 and 5 show that the student's perception of communication openness as one of the components of the knowledge sharing climate is able to predict 7.6 % of the variance in the level of using collaboration as a style for task conflict resolution.

H4B: Student's perception of the level of communication openness and the perception of knowledge sharing culture as components of knowledge sharing climate in the university affect significantly and positively his/her level of using competitive task conflict resolution style.

To test this hypothesis the researcher used regression test and found the following.

Table 6. Regression model for hypothesis 4B

Variables entered	R	R square	Adjusted R square	Std. Error of the estimate	F	Sig.
Perception of communication openness	0.419	0.175	0.159	1.49643	10.524	0.000

Dependent variable is competition conflict resolution style.

– The model is significant and the coefficients are as follows.

Table 7. Coefficients of regression model for H4B

Model	Unstandardized coefficients		Standardized coefficients	T	Sig.
	B	Std. Error	Beta		
Constant	1.298	0.816		1.590	0.115
Perception of communication openness	0.046	0.022	0.192	2.076	0.041
Perception of knowledge sharing culture	0.105	0.028	0.342	3.701	0.000

The Tables 6 and 7 show that the student's perception of communication openness and the perception of knowledge sharing culture are the components of the knowledge sharing climate that are able to predict almost 16 % of the variance in the level of using competition as a style for task conflict resolution.

H4C: Student's perception of the level of communication openness as one of the components of knowledge sharing climate in the university affects significantly and positively his/her level of using compromise task conflict resolution style.

To test this hypothesis the researcher used regression test and found the following.

Table 8. Regression model for hypothesis 4C

Variables entered	R	R square	Adjusted R square	Std. error of the estimate	F	Sig.
Perception of communication openness	0.321	0.103	0.094	1.78886	11.497	0.001

Dependent variable is compromise conflict resolution style.

– The model is significant and the coefficients are as follows.

Table 9. Coefficients of regression model for H4C

Model	Unstandardized coefficients		Standardized coefficients	T	Sig.
	B	Std. Error	Beta		
Constant	2.499	0.716		3.488	0.001
Perception of communication openness	0.114	0.034	0.321	3.391	0.001

The Table. 8 and 9 show that the student's perception of communication openness as a component of the knowledge sharing climate is able to predict 9.4 % of the variance in the level of using compromise as a style for task conflict resolution.

H4D: Student's perception of the level of knowledge sharing culture as one of the components of knowledge sharing climate in the university affects significantly and positively his/her level of using accommodation task conflict resolution style.

To test this hypothesis the researcher used regression test and found the following.

Table 10. Regression model for hypothesis 4D

Variables entered	R	R square	Adjusted R square	Std. error of the estimate	F	Sig.
Perception of knowledge sharing culture	0.332	0.111	0.102	1.76018	12.424	0.001

Dependent variable is accommodation conflict resolution style.

– The model is significant and the coefficients are as follows.

Table 11. Coefficients of regression model for H4D

Model	Unstandardized coefficients		Standardized coefficients	t	Sig.
	B	Std. Error	Beta		
Constant	2.499	0.716		3.488	0.001
Perception of knowledge sharing culture	0.114	0.034	0.321	3.391	0.001

The Table. 10 and 11 show that the student's perception of knowledge sharing climate as a component of the knowledge sharing climate is able to predict 10.2 % of the variance in the level of using accommodation as a style for task conflict resolution. As a conclusion after the analysis of hypothesis 4 we accept it.

H5: Knowledge Sharing Climate in general is significantly and positively affecting the highest assertiveness and collaboration style of conflict resolution which is collaboration conflict resolution style.

To test this hypothesis the researcher used regression test and found the following.

Table 12. Regression model for hypothesis 5

Variables entered	R	R square	Adjusted R square	Std. error of the estimate	F	Sig.
Perception of knowledge sharing culture	0.249	0.062	0.053	2.48066	6.597	0.012

Dependent variable is collaboration conflict resolution style.

– The model is significant and the coefficients are as follows.

Table 13. Coefficients of regression model for H5

Model	Unstandardized coefficients		Standardized coefficients	t	Sig.
	B	Std. Error	Beta		
Constant	5.969	1.596		3.740	0.000
Perception of knowledge sharing climate	0.062	0.024	0.249	2.568	0.012

The Tables 12 and 13 show that the student's perception of knowledge sharing climate is able to predict 5.3 % of the variance in the level of using collaboration as a style for task conflict resolution. According to the analysis of hypothesis 5 we can accept this hypothesis.

7 Research Model

Based on the hypotheses testing the following is a tested model for predicting the conflict resolution styles and which knowledge sharing climate variables can affect them in the case of this study (Fig. 1).

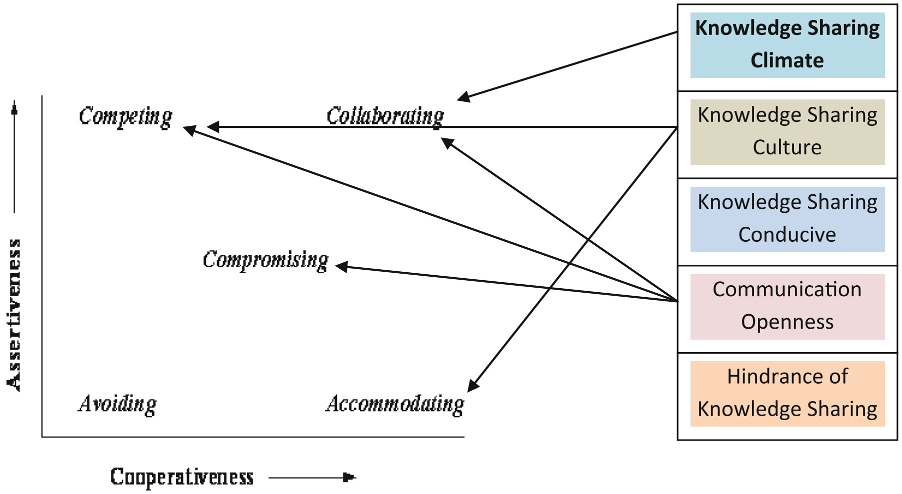


Fig. 1. Tested model of the prediction of conflict resolution styles by the variance in knowledge sharing climate

From the figure we can conclude that the knowledge sharing climate is capable to predict the middle levels to high levels of assertive and cooperative conflict resolution styles. Knowledge sharing climate is not able to predict the low levels of cooperativeness and assertiveness of conflict resolution which is represented by the avoidance style.

8 Research Conclusion and Recommendations

From the interviews, secondary data collected, and the questionnaire analysis we can conclude that the University of Sharjah is capable to predict significant part of the variance in cooperativeness and assertiveness in solving task conflicts among its students only if the level of cooperativeness and assertiveness are middle to high levels. If the student is avoiding task conflicts then the knowledge sharing climate is not predicting the low cooperativeness and the low assertiveness represented by avoidance style. It was found that the most effective component in the knowledge sharing climate is the communication openness because it affects positively three styles of conflict resolution and can predict significant ratio of the variance in the levels of cooperativeness and assertiveness in using those three styles. The next important component of the knowledge sharing climate is the perceived knowledge sharing culture which was mainly part of the Arabian culture respondents belong to. The perception of the knowledge sharing culture in the University of Sharjah is significantly predicting two opposing styles of conflict resolution which are the competitive conflict resolution style and the accommodation conflict resolution style. The first one represents high level of assertiveness and low level of collaboration. The second one represents high level of collaboration and low level of assertiveness. The research also found that the general perception of the knowledge sharing climate is positively related in the case of Sharjah University with collaboration style of conflict resolution. This means that the more encouragement of having healthy knowledge sharing climate we can afford, the higher levels of collaboration and assertiveness the conflict resolution we will get. This may lead to constructive conflict resolution innovations and will lead to having productive functional task conflicts among students rather than getting into dysfunctional task conflicts or avoiding all types of conflicts which means less learning. We recommend that researchers do the same research on larger sample and on more cases to reach a generalization of results. We also recommend that the University of Sharjah makes more effort in the aspect of rewarding students who have higher levels of knowledge sharing and are using collaborative conflict resolution style heavily and then to measure the effect of that on the innovation level and the productivity of the learning process in the university. The conclusion and recommendation of this research is true only regarding the sampling units of this case study. The time limitation also made responses for the survey limited to 102 respondents only. More respondents could have been included if more time was available. Another limitation is that the foreign students are not represented in the sample due to the fact that they are rare in the University of Sharjah and that the research scope focused on the knowledge sharing culture as part of the Arabian culture specifically. We suggest measuring knowledge donation's effect on conflict assertive solutions and collaborative solutions. We suggest measuring the effect of mixed cultural knowledge sharing perceptions on the project team's conflict levels. We also suggest finding other variables to explain the usage of conflict resolution styles among students in the implementation of their projects other that knowledge sharing climate in order to explain its variance and to control it in the future.

References

Boros, S., Meslec, N., Cursue, P., Emons, W.: Struggles for cooperation: conflict resolution strategies in multicultural groups. J. Manag. Psychol. **25**(5), 539–554 (2010)

Gupta, K.: A comparative analysis of knowledge sharing climate. Knowl. Process Manag. **15**(3), 186–195 (2008)

Han, G., Harms, B.: Team identification, trust, and conflict: a mediation model. J. Confl. Manag. **21**, 20–43 (2010)
Hooff, B., Ridder, J.: Knowledge sharing in context: the influence of organizational commitment, communication climate and CMC use on knowledge sharing. J. Knowl. Manag. **8**(6), 117–130 (2004)
Jehn, K.A.: A qualitative analysis of conflict types and dimensions in organizational groups. Adm. Sci. Q. **42**, 530–557 (1997)
Leung, A.S.: Interpersonal conflict and resolution strategies: an examination of Hong Kong employees. Team Perform. Manag. **14**(3/4), 165–178 (2008)
Lewicki, R.J., McAllister, D.J., Bies Bunker, R.J.: Trust and distrust: new relationships and realities. Acad. Manag. Rev. 23(3), 438–458 (1998) http://www.jstor.org/stable/259288
Lin, H.: Knowledge sharing and firm innovation capability: an empirical study. Int. J. Manpower **28**(¾), 315–332 (2007)
Lu, L., Zhou, F., Leung, K.: Effects of task and relationship conflicts on individual work behaviors. Int. J. Confl. Manag. **22**, 131–150 (2011)
Rousseau, D.M.: Why workers identify with organizations. J. Organ. Behav. 19(3), 217–233 (1998)
Thomas, K.W., Kilmann, R.H.: The Thomas-Kilmann Conflict Mode Instrument. XICOM Inc., Tuxedo (1974)
Yesilbas, L., Lombard, R.: Specification of a repository to support collaborative knowledge exchange in IPPOP project. Comput. Ind. **75**, 690–710 (2006)
Yesilbas, L., Lombard, M.: Towards a knowledge repository for collaborative design: focus on conflict management. Comput. Ind. **55**, 335–350 (2004)
www.sharjah.ac.ae
http://academic.engr.arizona.edu/vjohnson/ConflictManagementQuestionnaire/CMQResults.asp
http://www.surveymonkey.com/s.aspx?sm=879oL5LeM4vIkgOT8gx%2bZw%3d%3d–knowledge sharing questionnaire

Knowledge Sharing and Employee Development in Oil and Gas Companies in the United Arab Emirates

Moosa Alhanshi(✉) and Hussain Albraiki

British University in Dubai (BUID), 4627 Dubai, UAE
Alhanshi@gmail.com, hussain.albraiki@zadco.ae

Abstract. It is a fact nowadays that knowledge is the number one resource of organizations, and having that knowledge managed and shared effectively gives the organization the competitive advantage over other organizations. Many oil and gas organizations believe that having a good system and environment for knowledge sharing will have a positive effect on the learning curves of both the individual and the organization. In this study the researchers test the relationship between knowledge sharing and employee development and address the key knowledge-sharing factors that might have a positive relationship with employee effectiveness and subsequently their development. An empirical study was carried out to investigate this relationship in two oil and gas companies in the United Arab Emirates (UAE). One hundred and fifty questionnaires were distributed in both companies. The questionnaire consisted of 46 items addressing four factors of knowledge sharing: *Organizational culture, Individual Communications skills, Procedural justice,* and *Supervision and feedback*. The dependent variable is *employee development*. Of the responses receive, 124 were valid to use. The analysis was undertaken using the statistical package of social science software (SPSS) and the results showed a positive relationship between the global variables (knowledge sharing and employee development), but surprisingly one knowledge-sharing factor – Individual Communications skills - showed no significant relationship with employee development. At the end of the paper the researchers address the limitations of the study and make some recommendations to organizations on how to improve knowledge sharing which should lead to employee development and accordingly give the organization a better competitive advantage. Finally the researchers suggest some directions for future research.

Keywords: Knowledge sharing · Employee development · Petroleum industry

1 Introduction

Knowledge lies at the heart of today's economy and is one of the main competitive advantage factors for any organization [7]. The importance of the knowledge-sharing concept in elevating the organization's performance is well known and applied by the big oil and gas companies such as British Petroleum, Exxon Mobile and Chevron [11]. It is applied because those organizations need to survive in the face of today's hard competitive environment [7]. The employees of any organisation are the main drivers of the cycle

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 190–208, 2015.
DOI: 10.1007/978-3-319-21009-4_15

of the knowledge-sharing process and their effectiveness and willingness to cooperate is important to achieve the organization's goals and better performance [16]. Employee development and its related programs and methods are still a main target of any organisation [28] due to the resulting positive impact when applied and managed in a professional way [20]. A study by author [38] was conducted to investigate the relationship between employee work performance and their cooperation with others. The study concluded that such employees who showed an ability and interest to participate in knowledge-sharing activities gain better experience and add knowledge which increases their performance and accordingly affects their overall development in a positive way. So the employee development programs along with knowledge-sharing factors and techniques are proved to play a role in both employee and organisation performance and have become an important goal in the organization's strategy [12]. In general, the objective of organizations is to have a competitive set of unique services to be able to stand in today's economy. Knowledge management systems in oil and gas organizations have been used for two decades in different contexts and from different perspectives [11]. One application of the knowledge management system is knowledge sharing in which the individual should be identifying, acquiring, applying, creating, developing, preserving and measuring the knowledge of the organisation [41]. The process of sharing knowledge is considered the main way to create new knowledge between people [14]. The effect of such knowledge shared in an organisation has long been a concern within the academic area [6]. Employee development on the other hand has long attracted the interest of social science scholars and many studies such as studies by authors [16, 20, 37] have investigated in depth the importance and usage of different employee development approaches to enhance the organisation's and the employee's performance.

A deep look into the literature of knowledge sharing and its related factors on one side and employee development methods and approaches on the other side can conclude the existence of a positive relationship between them. Accordingly organizations that fail to provide different means for knowledge sharing will be affected. That affect in turn will reach the employee level, which will ultimately affect the overall organisation performance [42]. This study investigates the direct relationship between knowledge sharing and its factors with employee development in oil and gas organizations in the United Arab Emirates (UAE). The petroleum industry is the context of this study. UAE has the fourth-largest oil reserve in the world and the world's seventeenth-largest reserve of gas. Thirty per cent of the country's GDP comes from the petroleum sector [43]. This study is considered important for both the organizations as practitioners and for academic personnel. The study offers some recommendations that organisation and employees should ensure in order to increase performance. The study of the direct relationship between knowledge sharing and employee development had not been covered explicitly in previous research particularly in the UAE. Furthermore, this study is one of only a few that has investigated this relationship by focusing mainly on the employees of two major oil and gas companies in the UAE. The study consists of a literature review about knowledge sharing in general followed by a review of the factors reported in the literature that affect knowledge sharing. The literature of employee development related to knowledge sharing is also addressed. The methodology used is discussed, followed by findings and analysis. The study concludes with recommendations, limitations and suggestions for further research.

2 Literature

Due to the nature of the study, the researchers split the literature review into two main parts. The first concentrates on knowledge sharing, how it is developed, and its main features. In addition it gives details about previous studies which addressed the factors that affect sharing knowledge. The researchers chose four factors that directly affect sharing of knowledge based on previous studies. The second part addresses some studies on employee development and how that is related to knowledge sharing and knowledge-sharing factors.

2.1 Knowledge Sharing

Knowledge by nature is either *tacit* or *explicit* [11]. Tacit knowledge is the 'know-how' knowledge which comes through the experience of each individual [26]. The explicit form of knowledge is the hard copy type of knowledge and it can be transferred and stored and so shared easily with others [46]. Tacit knowledge is hard to handle and stored and in fact is hard to identify, but on the other hand it plays an important role in problem solving and time saving [44], and, if managed and shared effectively, can significantly enhance employee performance [34]. One of the more effective ways of creating new knowledge is encouraging a knowledge-sharing process among people [44, 34]. It is one of the main systems in the knowledge management concept which directly affects the organisation and the employee [44]. Knowledge sharing within an organisation had been looked as a contribution of an individual to the collective pool of corporate knowledge [26]. Knowledge sharing has long been recognized as an important research topic [45]. Knowledge sharing –also called knowledge *diffusion* - can be implemented successfully depending on the level of knowledge between knowledge seekers and owners [41]. Implementation of such a system in any organisation will require the understanding of the concept and the barriers involved [8] and then motivate the employees in the organisation to share the knowledge [46]. There is no unique definition for knowledge sharing, but it can be defined as an approach in identifying, acquiring, applying, creating, developing, preserving and measuring the knowledge of the organisation [14]. Many aspects need to be known when addressing the knowledge-sharing concept such as the factors influencing knowledge sharing which can be summarized as Organisational culture [19], Individual Communications skills [3], Procedural justice [2], and Supervision and feedback [37]. The fact that tacit knowledge is based on personal experience makes it a main part that could be connecting knowledge sharing and employee development where employees can enhance their work performance and accordingly increase the chance for better development when they cooperate and share knowledge with other colleagues [38].The four factors mentioned above have a direct influence on knowledge sharing and it has also been proved that they directly improve organisation performance and employee development [16]. Based on this, we set out our main hypothesis:

H1: There is a positive relationship between knowledge sharing and employee development.

2.1.1 Organizational Culture

Understanding the definition of organisational culture is important if we are to understand its relationship to knowledge sharing. Although there is no unique definition for organisational culture, for the purposes of our study, the definition by author [30] seems to be suitable, he define organisational culture as the knowledgeable members of a group and their willingness to share knowledge. It actually gives the employees day-to-day guidance on how to communicate with each other [1]. Organisational culture could be considered as a key enabler for sharing knowledge [5, 19]; in a way that the associated norms and practices that come with organisational culture elevate the team's work style which in turn affects them positively and encourages knowledge sharing [8]. The importance of organisational culture on knowledge sharing is that when such a system is adopted in an organisation on a daily basis, then the employees will not feel as if they are being forced to share knowledge; instead they consider it as a part of their daily tasks [24]. Furthermore the organisational culture affects - both directly and indirectly - the individual in the organisation and accordingly helps them to develop their skills and so improve their performance [13]. Therefore:

H2: There is a positive relationship between organizational culture as a knowledge-sharing factor and employee development.

2.1.2 Procedural Justice

The idea of social justice has long been a subject of interest for social psychologists because of its effect and importance [33]. As a theory, procedural justice "suggests that the perceived fairness of distributed outcomes should be determined by the procedures of distribution of outcomes to workers" [39]. This is the feeling of the individual that the decision-making process and outcome are fair; which means that subsequent behavior and attitude of employees will be affected by justice. It is a fact that people's ability and willingness to cooperate with other people is affected by justice [32]. One form of that cooperation is the people's ability and willingness to share knowledge.

Procedural justice has an influence on and plays a key role in knowledge sharing of personnel; which in turn affects the knowledge at organizational and senior management level [35]. An example of that is when an organization undergoes a change that needs employees' commitment, which can be elevated and increased with the presence of procedural justice [23]. In general, having a fair process of decision making and other management roles will ensure trust between employees and the organization which should accordingly encourage personnel to share knowledge and communicate more effectively [2]. The role of procedural justice not only influences the cooperation and commitment of employees but also affects their performance in many ways; one such way is the employee performance appraisal [31]. The performance appraisal involves tools used by organizations in many activities such as raising employee's performance [10]. So procedural justice has a direct effect on employee development, and managers should take it into consideration when taking the decision to promote employees and apply a fair employment law. Failing to do so could affect the employee's behavior and accordingly may force them to quit [36]; accordingly:

H3: there is a positive relationship between procedural justice as a knowledge-sharing factor and employee development.

2.1.3 Individual Communication Skills

In order for knowledge sharing to take place, employees should be equipped with certain skills in order to effectively use and share knowledge [22]. Such skills are like the ability to share and the motivation to give knowledge [3]. In addition, some experience will also be needed which help to accelerate the sharing will process because, when there are senior personnel who are experts on some fields of knowledge, they tend to train and share their knowledge with junior individuals [9]. The above-mentioned skills prove the positive relationship with knowledge sharing, and on the other hand, having good communication skills directly affects the employee's performance. Many organizations used such skills to evaluate employees' overall development and identify weaknesses in order to find ways of improvement to elevate employee excellence. This will eventually lead to customer satisfaction and so achieve the organization's goals [37]. In addition to the individual skills there should be an effective fit between the organisation culture, the environment, and the employee's skills and characteristics; such a relationship will lead to more commitment and accordingly more effective knowledge sharing [15]. Furthermore, it is considered very important for an individual to be equipped with new technology methods that are available nowadays [18]. Technological skills such as using emails, instant messaging software, sharing sites and other smart applications are now considered the very basics for employees to have; those methods could help make the sharing of knowledge more effective and give the organisation an advantage against their competitors [11]. In general, the new technology methods gained by employees are directly affecting their ability to share knowledge and their chance to raise performance [40]. Organizations need to give more attention to the new communication tools used to share knowledge to raise the performance of both the organisation and employees [11]. Based on this,

H4: There is a positive relationship between individual communication skills as a knowledge-sharing factor and employee development.

2.1.4 Supervision and Feedback

Knowledge sharing by its very nature is a relation between seekers and receivers of knowledge [14]. In many organizations, such sharing of knowledge should be associated with good supervision feedback; that feedback should be in the form of a close relationship between the supervisor and subordinates [4]. The professional relationship between employees and their supervisors is important to improving the employee's level of understanding using (for example) a report review which should include a feedback about the employee's work [25]. That kind of feedback would enhance the learning and knowledge of employees which raises their performance [17]; [21], this forms the basis of the next hypothesis.

H5: There is a positive relationship between supervision and feedback as a knowledge-sharing factor and employee development.

2.2 Employee Development

Employee development in its normal form can be defined as assisting employees to learn more in order to perform better but such a view needs to be looked at and added to it in order to understand "the wide range of conditions within which employee development programs exist in organizations" [16]. There are many employee development methods that are proven good and used by organisation to improve employee performance and organisational outcomes and which at the same time are directly related to the concept of knowledge sharing such as "off-the-job and on-the-job training programs, educational programs and seminars, job rotations, and mentoring programs" [16]. Many studies have been conducted in the area of employees' development and its effect on employee and organisation performance. For example at employee level; author [20] study the influence of different learning experience and job performance which shows positive relationship. The organisation should help employees by providing different learning chances to make them perform more and, accordingly, that would lead to better performance. One of such major learning chances is the knowledge and learning that employees get because of their cooperation and communication with other employees [37]. This proves the relationship between the knowledge sharing and employee development.

2.3 Conceptual Framework

Figure 1 demonstrates the hypotheses relationship between independent and dependent variables.

3 Research Methodology

The method used in this study is a quantitative approach through the administering of a questionnaire which has been proved suitable for such research, as this method generates accurate and reliable data using a different range of samples and many proved tests to check the significance of the relations. This section explains the methods and procedures that researchers use with the instrument, the variable, and measures applied.

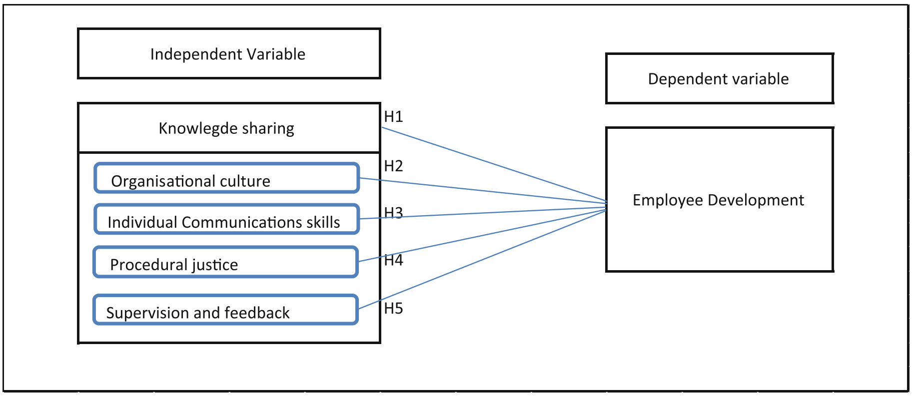


Fig. 1. Conceptual framework

3.1 Samples and Procedures

The context of the study as mentioned above is the petroleum industry in the UAE. The study was conducted in two major oil organizations in the UAE; one in Dubai and the other in Abu Dhabi. Self-administered questionnaires were distributed (online and hard copy) in order to investigate and explain the addressed hypotheses and check the significance of the relationship between the dependent and independent variables as per the proposed conceptual framework above. One hundred and fifty employees from both organizations were selected; all holding full-time contracts and the majority of whom are engineers from different levels of management. Out of 150 questionnaires 132 responses come back, and out of those 124 questionnaires were valid to be used in the analysis with a response percentage of 93 %. Each questionnaire (online and hard copy) had, attached, an introduction explaining what the questionnaire is about and what the purpose of the study.

3.2 Study Instruments

A questionnaire survey consisting of 46 questions was distributed by hand and online link through emails and smart phones. The survey was divided into three sections: the *demographic* section which contained nine items; the *knowledge-sharing* section divided into four factors with 19 questions; and the *employee development* section containing 26 questions. A pilot study with 25 responses was tested at the start of the study to make sure that the factors were loaded well and that items were consistent. After acquiring the results of the pilot test, the researchers had the confidence to start the actual work.

3.3 Variables and Measures

The study contains three main variables:

Demographic Variable: it consists of nine areas: gender, age, marital status, nationality, education, organization type, job status, number of years in organization and number of years on current job. The researchers used a scale ranging from two (e.g. gender) to six (e.g. education level). The scale is developed by [29].

Knowledge-Sharing Variable: knowledge sharing is a multidimensional variable consisting of four factors, comprising a total of 24 questions. Those factors are organisational culture (five questions), individual communication skills (five questions), procedural justice (six questions) and supervision and feedback (five questions). The 5-point Likert format is used in this study. The respondents were asked to choose from the following responses: 1 for “strongly agree”, 2 for “agree”, 3 for “Neutral”, 4 for “disagree” and 5 for “strongly disagree”.

Employee Development: employee development is the one-dimensional variable and used the same scale used for the knowledge-sharing variable.

4 Data Analysis and Discussion

This part of the study introduces the analysis of the data and the findings that were reached based on the results of articulated tests. Furthermore, discussions and comments are presented later. In order to analyses the data that were gathered for the study, the Statistical Package for the Social Sciences (SPSS) was principally employed to do all the required tests to check the hypotheses. Table 1 Factors Influencing on Privacy Preferences in Healthcare Environment.

Table 1. Demographic data

Demographic Variables	Gender	Age	Marital Status	Nationality	Education	Organisational Type	Job status	Number of years in Current organisation	Number of years in Current Position
Male Female	75 (60.5%) 49 (39.5%)								
less than 25 25-35 36-46 47-57 58 or above		26 (21.0%) 69 (55.6%) 22 (17.7%) 5 (4.0%) 2 (1.6%)							
Married Not married			87 (70.2%) 37 (29.8%)						
UAE national Non-UAE national				45 (36.3%) 78 (62.9%)					
Less than high School High School Diploma Bachelors Masters PhD					2 (1.6%) 8 (6.5%) 40 (32.3%) 60 (48.4%) 13 (10.5%) 1 (.8%)				
Private Governmental Semi-Governmental						50 (40.3%) 29 (23.4%) 45 (36.3%)			
First Level Middle Level Lower Level							33 (26.6%) 58 (46.8%) 33 (26.6%)		
a year or less 2-7 years 8-13 years 14-19 years 20 years or above								12 (9.7%) 75 (60.5%) 26 (21.0%) 8 (6.5%) 3 (2.4%)	
a year or less 2-7 years 8-13 years 14-19 years 20 years or above									41 (33.1%) 77 (62.1%) 4 (3.2%) 1 (0.8%) 1 (0.8%)

4.1 Descriptive Statistic

In this section, the analyzed data are looked at in an abstract way from a descriptive statistics point of view, the aim of which is to illustrate the vital structures and main features of the data in the study. Table 1 below depicts the results of demographic data analysis.

As shown in Table 1 above, the demographic findings and career variables were presented in a simply summarized format for the study sample. In the study, the number of male participants was higher than female participants, representing a 60.5 % to 39.5 % dominant-male to female ratio. Furthermore, the majority of the participants fell in the middle age group of (25–35) years representing 55.6 % of the total sample. The number of married respondents outnumbered the unmarried respondents with a ratio of 70.2 % to 29.8 %. In terms of nationality of respondents, UAE nationals represented 36.3 % of the respondents, while Non-UAE nationals were greater, comprising 62.9 % of the sample. The majority of the study sample showed a high level of education, with respondents holding graduate and postgraduate degrees. The respondents with Bachelors and High Diploma degrees collectively represented more than the half of the study sample −48.4 % and 32.3 %, respectively. In addition, the study sample was comprised of governmental, semi-governmental and private sectors, with a noticeable difference between them, at 23.4 %, 36.3 %, and 40.3 %, respectively. The respondents varied in their career level, seen as being with equal numbers at entry level and first level, representing 26.6 % each, and collectively representing more than half of the study sample. On the other hand, employees at the middle level of their career represented 46.8 % of the study sample. The sample data show that the majority of employees had work experience that ranged between two and seven years with their current employers at their current position, accounting for 60.5 % and 62.1 % respectively. However, this varied among other respondents, showing the least who had more than 15 and 20 years of experience in the same organization and worked at same position, with one responded (0.8 %) to each respectively.

4.2 Reliability Test

This part of the study examines the reliability of the administered research survey questionnaire. Following the study by author [29], the minimum acceptable value for reliability is 0.6. Table 2 below indicates the summary of reliability test.

From Table 2 above, the studies overall Cronbach's alpha value was 0.826; knowledge sharing was found to be 0 which is considered highly reliable, while employee development is 0. None of the items were excluded during the reliability texts.

4.3 Factor Analysis

In this section, the data ware assessed dimensionally in order to ensure that proper loading of the studied items under the factors. According to author [29] a value of 0.50 or above is set as the item's acceptance criteria in order to have a meaningful factors

Table 2. Reliability test

Variable	Cranach's alpha	# of items	# of items after deleting
Global: knowledge sharing	0.826	20	18
Global factor: organisational culture	0.819	5	5
Global factor: individual communications skills	0.831	5	4
Global factor: procedural justice	0.808	5	5
Global factor: supervision and feedback	0.813	5	4
Global: employee development	0.831	26	18
Overall	0.888	46	36

arrangement by improving scale reliability and excluding weakly loaded items from the subsequent analysis. Table 3 below summarizes the factor analysis results.

At the very beginning of the factor analysis, not all of the knowledge-sharing factor items were found to satisfy the abovementioned criteria. Each of the four factors had some poorly loaded items (with values less than 0.50). Thus, they were removed and the items were recomputed, yielding the current one presented in Table 2 above, only retaining the one with a value of 0.5 and above (shown in Table 3). Most of the items in the study showed proper loading. Hence, the analysis is considered done when the current items will are properly loaded under each factor. Table 3 below explains and provides the numerical data of factor items that are loaded.

Table 3. Factor loading for knowledge sharing

Rotated component matrix				
	Component			
	1	2	3	4
Item-1 organizational culture	0.81			
Item-2 organizational culture	0.655			
Item-3 organizational culture	0.606			
Item-4 organizational culture	0.837			
Item-1 supervision and feedback		0.764		
Item-2 supervision and feedback		0.793		
Item-3 supervision and feedback		0.742		
Item-1 procedural justice			0.587	
Item-4 procedural justice			0.526	
Item-5 procedural justice			0.76	
Item-6 procedural justice			0.591	
Item-4 individual communications skills				0.717
Item-5 individual communications skills				0.789

4.4 Correlation Test

In this section, the data are assessed and tested in order to investigate the relationship and its significance among the variables of the study. Table 4 below summarizes the correlation test results.

Table 4. Correlation test results

Correlations

		Organizational Culture	Individual Communications Skills	Procedural Justice and Trust	Supervision and Feedback	Knowledge Sharing Global	Employee Development Global
Organizational Culture	Pearson Correlation	1					
	Sig. (2-tailed)						
	N	51					
Individual Communications Skills	Pearson Correlation	.179	1				
	Sig. (2-tailed)	.209					
	N	51	51				
Procedural Justice and Trust	Pearson Correlation	.692**	.141	1			
	Sig. (2-tailed)	.000	.324				
	N	51	51	52			
Supervision and Feedback	Pearson Correlation	.412**	.053	.613**	1		
	Sig. (2-tailed)	.003	.714	.000			
	N	51	51	52	52		
Knowledge Sharing Global	Pearson Correlation	.759**	.337*	.878**	.763**	1	
	Sig. (2-tailed)	.000	.017	.000	.000		
	N	50	50	50	50	50	
Employee Development Global	Pearson Correlation	.381**	.130	.555**	.570**	.586**	1
	Sig. (2-tailed)	.006	.368	.000	.000	.000	
	N	50	50	51	51	50	51

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

In the table above, the correlation test, it is clearly shown that there is a positive relationship between knowledge sharing comprising its four studied factors and employee development. Furthermore, this relationship is found significant among the majority of the factors. All the factors of knowledge sharing are positively related to each other. This relationship is significant between the organizational culture factor and the procedural justice and trust factor, and the supervision and feedback factor, with the Pearson correlation values of 0.692 and 0.412, respectively. Likewise, they have significance values of 0.000 and 0.003, respectively, significant at the 0.01 level. Meanwhile, the procedural justice and trust factor and supervision and feedback factor are significant with the Pearson correlation value of 0.613 and significance value of 0.000, significant at the 0.01 level. Moreover, it appears that these factors have a significant relationship with the global factor. Knowledge sharing as a global variable shares a positively significant relationship with its inclusive factors, namely the organizational culture, individual communication skills, procedural justice and trust, and supervision and feedback as well as the global variable of employee development, with the values of Pearson correlation being 0.759, 0.337 0.878, 0.763 and 0.586, respectively.

In addition, they all shared the significance value of 0.000 for all relationships, significant at the 0.01 level, except for the individual communication skill factor which has a significance value of 0.017 and is significant at the 0.05 level. Furthermore, these factors show a positive relationship with employee development. However, it is only significant with the organisational culture factor and the procedural justice and trust factor, and supervision and feedback factor with the Pearson correlation values of 0.381, 0.555 and 0.570, respectively; likewise, they have significance values of 0.006, 0.000 and 0.000, respectively, significant at the 0.01 level. As seen from the above, most of the variables were significant at the 0.01 level, which means that there is a probability that 99 % of the variables will have the same relationship if the study is conducted again. Finally, and as expected from the literature review, the results succeeded in showing a positive significant relationship between the main factors of knowledge sharing and extended this relationship to the employee development, with some limitations, although still supporting hypotheses H1, H2, H3 and H5.

4.5 Discussion of Main Findings

In this section, the data results are discussed and commented on to confirm the findings and finalize the decisions on the proposed hypotheses. The results of the correlation and regression tests are mainly used to assess the relationships among the research hypotheses variables and factors. H1: There is a positive relationship between knowledge sharing and employee development. Examining this hypothesis's validity, correlation results were looked at, where the R-value is 0.586 and is significant (Sig. Level = 0.000). This refers to statistical evidence on the existence of a relationship between the knowledge-sharing practice by employees in an organisation and the level of employee development created in this organisation. Regression test was used to further assess this relationship, and the results are shown in Table 5.

From the above table, it is clearly seen that the R-value is 0.586, R-square value is 0.343, adjusted R-square value is 0.330, and F-value is 25.103 with high significance (0.000). Also, it is noted that the beta weight of the knowledge-sharing variable is 0.586, and the t-value is 5.01, which is significant (Sig. Level = 0.000). These results imply that employee development is significantly related to knowledge sharing, and it could be concluded that hypothesis 1 (H1) is upheld this conclusion is consistent with previous studies stating that knowledge sharing has a direct effect on the organisation and the employee development [47].

Table 5. Regression test between knowledge sharing and employee development

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.586[a]	.343	.330	9.104

a. Predictors: (Constant), Knowledge Sharing Global Factor

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2080.711	1	2080.711	25.103	.000[b]
	Residual	3978.569	48	82.887		
	Total	6059.280	49			

a. Dependent Variable: Employee Development Global Factor
b. Predictors: (Constant), Knowledge Sharing Global Factor

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	13.947	9.503		1.468	.149
	Knowledge Sharing Global Factor	.637	.127	.586	5.010	.000

a. Dependent Variable: Employee Development Global Factor

H2: There is a positive relationship between organizational culture as a knowledge-sharing factor and employee development. To test the validity of this hypothesis, correlation results were looked at, where the R-value is 0.381 and is significant (Sig. Level = 0.006). This refers to a relationship between knowledge-sharing culture practiced in an organization and the level of employee development created in this organization. A regression test was used to further assess this relationship, and the results are shown in Table 6. From the above, it is clearly seen that the R-value is 0.381, R-square value is 0.145, adjusted R-square value is 0.127, and F-value is 8.132 with high significance (0.006). Also, it is noted that the beta weight of the knowledge-sharing variable is 0.381, and the t-value is 2.852, which is significant (Sig. Level = 0.006). These results imply that employee development is significantly related to positive organisational culture in terms of knowledge-sharing practice, and it could be concluded that hypothesis 2 (H2) is established. This finding confirms previous studies in the field [47]; that organisational culture as a knowledge-sharing factor improves employee development, and that organisational culture has a direct effect on the employees and their development [13]. H3: There is a positive relationship between procedural justice as a knowledge-sharing factor and employee development.

To test the validity of this hypothesis, correlation results were looked at, where the R-value is 0.555 and is significant (Sig. Level = 0.000). This refers to a relationship between procedural justice as a knowledge-sharing factor in an organisation and the level of employee development created in this organisation. A regression test was used to further assess this relationship, and the results are shown in Table 7.

Table 6. Regression test

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.381[a]	.145	.127	10.390

a. Predictors: (Constant), Organizational Culture Factor

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	877.796	1	877.796	8.132	.006[b]
	Residual	5181.484	48	107.948		
	Total	6059.280	49			

a. Dependent Variable: Employee Development Global Factor
b. Predictors: (Constant), Organizational Culture Factor

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	39.897	7.586		5.259	.000
	Organizational Culture Factor	1.698	.595	.381	2.852	.006

a. Dependent Variable: Employee Development Global Factor

Table 7. Regression test between procedural justice as knowledge sharing factor and employee development

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.555[a]	.308	.294	9.254

a. Predictors: (Constant), Procedural justice and trust Factor

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1866.508	1	1866.508	21.795	.000[b]
	Residual	4196.237	49	85.637		
	Total	6062.745	50			

a. Dependent Variable: Employee Development Global Factor
b. Predictors: (Constant), Procedural justice and trust Factor

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	34.554	5.844		5.913	.000
	Procedural justice and trust Factor	1.733	.371	.555	4.669	.000

a. Dependent Variable: Employee Development Global Factor

From the above, it is clearly seen that the R-value is 0.555, R-square value is 0.308, adjusted R-square value is 0.294, and F-value is 21.795 with high significance (0.000). Also, it is noted that the beta weight of the knowledge-sharing variable is 0.555, and the t-value of is 4.669, which is significant (Sig. Level = 0.000). These results imply that employee development is significantly related to procedural justice and trust of the knowledge-sharing practice in an organisation, and it could be concluded that hypothesis 3 (H3) is established. This finding confirms those of previous studies in the field [47]; that procedural justice as a knowledge-sharing factor improves employee development, and procedural justice plays an important role in influencing the cooperation and commitment of employees and their performance in many ways [31].

H4: There is a positive relationship between individual communication skills as a knowledge-sharing factor and employee development.

Examining this hypothesis's validity, correlation results were looked at, where the R-value is 0.130 and is not significant (Sig. Level = 0.368), which means that individual communication skills supports the knowledge-sharing practice but does not affect the level of employee development in the organisation. From here it could be concluded that hypothesis 4 (H4) is not established. Unexpectedly, this did not comply with previous studies that stated that gaining good technological capability in communication tools and techniques [40] and gaining more modern ones [11] will affect employees directly in their ability to share knowledge and in their development. The reason for the change might be the recent advances in communication technology that have removed all barriers to knowledge sharing, thus affecting employees' development. H5: There is a positive relationship between supervision and feedback as a knowledge-sharing factor and employee development.

To test the validity of this hypothesis, correlation results were looked at, where the R-value is 0.57 and is significant (Sig. Level = 0.000). This refers to a relationship between supervision and feedback as a knowledge-sharing factor in an organisation and the level of employee development created in this organisation. A regression test was used to further assess this relationship, and the results are shown in Table 8.

From the above table, it is clearly seen that the R-value is 0.57, R-square value is 0.325, adjusted R-square value is 0.311, and F-value is 23.58 with high significance (0.000). Also, it is noted that the beta weight of the knowledge-sharing variable is 0.57, and the t-value is 4.856, which is significant (Sig. Level = 0.000). These results imply that employee development is significantly related to supervision and feedback of the knowledge-sharing practice in an organisation, and it could be concluded that hypothesis 5 (H5) is established. This finding confirms previous studies in the field [47]; that supervision and feedback as a knowledge-sharing factor improves employee development, and supervision and feedback enhances employees' knowledge [34] which in turn leads to their development [17, 21].

Table 8. Regression test between supervision and feedback as knowledge factor sharing and employee development

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.570[a]	.325	.311	9.140

a. Predictors: (Constant), Supervison and Feedback Factor

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1969.676	1	1969.676	23.580	.000[b]
	Residual	4093.069	49	83.532		
	Total	6062.745	50			

a. Dependent Variable: Employee Development Global Factor

b. Predictors: (Constant), Supervison and Feedback Factor

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	36.929	5.151		7.169	.000
	Supervison and Feedback Factor	2.263	.466	.570	4.856	.000

a. Dependent Variable: Employee Development Global Factor

5 Conclusion

This study investigated the relationship between knowledge-sharing practice and the employees' development in organizations. The results of the study were generally consistent with previous research, and confirm past studies. It depicted the correlation between knowledge sharing, and factors like organisational culture and procedural justice and trust, supervision and feedback, individual communication skills, and employee development measurement. All the factors showed positive relationships with employee development; they were also highly significant except for the factor "individual communication skills", which opposed previous findings. Some recommendations are listed below along with future research suggestions.

5.1 Recommendations

According to the findings from this study and aligned with the literature review, the researchers have a list of recommendations for the organizations and employees to follow in order to enhance the effectiveness of both to gain competitive advantage for the organisation and development for the employee: The researchers recommend introducing a new idea for sharing knowledge called the "organisational expert map" in which each organisation assigns one or more experienced personnel in each field inside the organisation, ensuring that they are accessible to employees to share knowledge, each according to his or her specialization. The organisation needs to evaluate the role of justice and evaluate the employees according to their performance, so it is recommended to include their cooperation and willingness to share knowledge in their performance appraisals. The organisation should ensure that the feedback environment is a common practice between supervisors and subordinates and it is recommended to conduct a bi-weekly meeting between both to check employees' progress. After this, the supervisor should communicate his feedback first to the employees and then to upper managers to either appreciate the good work or recommend more development programs such as training.

5.2 Limitations and Future Research Suggestion

There are always limitations in research in the academic area, and this study is no exception. In this study we have two main limitations; first the number of factors studied for knowledge sharing was limited to four because adding more factors would require significantly more time and effort. The second limitation was the number of respondents (124) and although this was suitable for this study, more respondents are always recommended. From the above limitations, the researchers have some suggestions for future studies on the same subject; one suggestion is to add more factors that affect knowledge sharing which are not addressed in this study, such as a reward system and information technology. To assist future researchers in this, the authors of this study have introduced those factors along with their questions. The second suggestion could be to review the factor that has no significant value - "individual communication skills" - and add more items, in addition to having more respondents for the questionnaire so that further analysis will either validate this study's findings or find positive relations between that factor and employee development. Finally to address the factor of communication skills, it could be helpful to look broadly in the organizational communication [27] as a contributor to that factor which may lead to better understanding.

References

1. Alavi, M., Kayworth, T., Leidner, D.: An empirical examination of the influence of organizational culture on knowledge management practices. J. Manag. Inf. Syst. **22**(3), 191–224 (2006)
2. Barney, J., Hansen, M.: Trustworthiness as a source of competitive advantage. Strateg. Manag. J. **15**(1), 175–190 (1994)
3. Beck, P., Solomon, I.: Experience, expertise and expert-performance research in public accounting. [Urbana-Champaign]: College of Commerce and Business Administration,

Bureau of Economic and Business Research, University of Illinois Urbana-Champaign (1989)
4. Brazel, J., Agoglia, C., Hatfield, R.: Electronic versus face-to-face review: the effects of alternative forms of review on auditors' performance. Acc. Rev. **79**(4), 949–966 (2004)
5. Brown, A., Starkey, K.: The effect of organizational culture on communication and information. J. Manag. Stud. **31**(6), 807–828 (1994)
6. Camelo-Ordaz, C., García-Cruz, J., Sousa-Ginel, E., Valle-Cabrera, R.: The influence of human resource management on knowledge sharing and innovation in Spain: the mediating role of affective commitment. Int. J. Hum. Resour. Manag. **22**(7), 1442–1463 (2011)
7. Carrillo, P.: Managing knowledge: lessons from the oil and gas sector. Constr. Manag. Econ. **22**(6), 631–642 (2004)
8. De Long, D., Fahey, L.: Diagnosing cultural barriers to knowledge management. Acad. Manag. Perspect. **14**(4), 113–127 (2000)
9. Fargher, N., Mayorga, D., Trotman, K.: A field-based analysis of audit workpaper review. Auditing J. Pract. Theor. **24**(2), 85–110 (2005)
10. Fletcher, C.: Performance appraisal and management: the developing research agenda. J. Occup. Organ. Psychol. **74**(4), 473–487 (2001)
11. Grant, R.: The development of knowledge management in the oil and gas industry. Universia Bus. Rev. **40**, 92–125 (2013)
12. Grant, R.: Toward a knowledge-based theory of the firm. Strateg. Manag. J. **17**(2), 109–122 (1996)
13. Gul, S., Khan, M., Razzaq, N., Saif, N., Ahmed, U., Ahmed, M.: Revisiting the corporate culture of pakistan: the rites and rituals of corporate life and employee performance. Ind. Eng. Lett. **2**(8), 1–10 (2012)
14. Hsiao, R.: Knowledge sharing in a global professional service firm. MIS Q. Executive **7**(3), 123–137 (2008)
15. Hyatt, T., Prawitt, D.: Does congruence between audit structure and auditors' locus of control affect job performance? Acc. Rev. **76**(2), 263–274 (2001)
16. Jacobs, R., Washington, C.: Employee development and organizational performance: a review of literature and directions for future research. Hum. Resour. Dev. Int. **6**(3), 343–354 (2003)
17. Jordan, A., Audia, P.: Self-enhancement and learning from performance feedback. Acad. Manag. Rev. **37**(2), 211–231 (2012)
18. KasapoUglu-Akyol, P.: Using educational technology tools to improve language and communication skills of ESL students. Novitas-ROYAL **4**(2), 225–241 (2010)
19. Keyton, J.: Communication and Organizational Culture. Sage Publications, Thousand Oaks (2005)
20. LaFleur, T., Hyten, C.: Improving the quality of hotel banquet staff performance. J. Organ. Behav. Manag. **15**(1–2), 69–93 (1995)
21. Latham, G., Ford, R., Tzabbar, D.: Enhancing employee and organizational performance through coaching based on mystery shopper feedback: a quasi-experimental study. Hum. Resour. Manag. **51**(2), 213–229 (2012)
22. Mckenzie, J., Van Winkelen, C.: Creating successful partnerships the importance of knowledge sharing. J. Gen. Manag. **31**(4), 55 (2006)
23. Michel, A., Stegmaier, R., Sonntag, K.: I scratch your back – you scratch mine. do procedural justice and organizational identification matter for employees' cooperation during change? J. Change Manag. **10**(1), 41–59 (2010)
24. Zamantili Nayir, D., Uzunçarşili, Ü.: A cultural perspective on knowledge management: the success story of Sarkuysan company. J. Knowl. Manag. **12**(2), 141–155 (2008)

25. Peecher, M.: A discussion of "audit review: managers' interpersonal expectations and conduct of the review". Contemp. Acc. Res. **19**(3), 445–448 (2002)
26. Reychav, I., Weisberg, J.: Good for workers, good for companies: how knowledge sharing benefits individual employees. Knowl. Process Mgmt. **16**(4), 186–197 (2009)
27. Roberts, K., O'Reilly, C.: Measuring organizational communication. J. Appl. Psychol. **59** (3), 321–326 (1974)
28. Seibert, S., Kraimer, M., Liden, R.: A social capital theory of career success. Acad. Manag. J. **44**(2), 219–237 (2001)
29. Suliman, A.: Work performance: is it one thing or many things? The multidimensionality of performance in a Middle Eastern context. Int. J. Hum. Resour. Manag. **12**(6), 1049–1061 (2001)
30. Van Maanen, J.: Tales of the Field. University of Chicago Press, Chicago (1988)
31. Tuytens, M., Devos, G.: The effect of procedural justice in the relationship between charismatic leadership and feedback reactions in performance appraisal. Int. J. Hum. Resour. Manag. **23**(15), 3047–3062 (2012)
32. Tyler, T., Blader, S.: Cooperation in Groups. Psychology Press, Philadelphia (2000)
33. Tyler, T.: Procedural justice, identity and deference to the law: what shapes rule-following in a period of transition? Aust. J. Psychol. **61**(1), 32–39 (2009)
34. Vera-Muñoz, S., Ho, J., Chow, C.: Enhancing knowledge sharing in public accounting firms. Acc. Horiz. **20**(2), 133–155 (2006)
35. Verbeke, A., Bachor, V., Nguyen, B.: Procedural justice, not absorptive capacity, matters in multinational enterprise ICT transfers. Manag. Int. Rev. **53**(4), 535–554 (2013)
36. Wan, H., Sulaiman, M., Omar, A.: Procedural justice in promotion decisions of managerial staff in Malaysia. Asia Pac. Bus. Rev. **18**(1), 99–121 (2012)
37. Wiele, T., Hesselink, M., Iwaarden, J.: Mystery shopping: a tool to develop insight into customer service provision. Total Qual. Manag. Bus. Excellence **16**(4), 529–541 (2005)
38. Wu, W., Yeh, R., Hung, H.: Knowledge sharing and work performance: a network perspective. Soc. Behav. Pers. Int. J. **40**(7), 1113–1120 (2012)
39. Yamaguchi, I.: Interpersonal communication tactics and procedural justice for uncertainty management of Japanese workers. J. Bus. Commun. **42**(2), 168–194 (2005)
40. Zhang, X., Venkatesh, V.: Explaining employee job performance: the role off online and offline workplace communication networks. MIS Q. **37**(3), 695–722 (2013)
41. Alavi, M., Leidner, D.: Review: knowledge management and knowledge management systems: conceptual foundations and research issues. MIS Q. **25**(1), 107 (2001)
42. Torraco, R., Swanson, R.: The strategic roles of human resource development. Hum. Resour. Plann. **18**, 10–21 (1995)
43. Wikipedia. Economy of the United Arab Emirates (2015). http://en.wikipedia.org/wiki/Economy_of_the_United_Arab_Emirates. Accessed 07 Feb 2015
44. Groff, T., Jones, T.: Introduction to Knowledge Management. Butterworth-Heinemann, Amsterdam (2003)
45. Cabrera, E., Cabrera, A.: Fostering knowledge sharing through people management practices. Int. J. Hum. Resour. Manag. **16**(5), 720–735 (2005)
46. Becerra-Fernandez, I., Sabherwal, R.: Knowledge Management. M.E. Sharpe, Armonk (2010)
47. Cho, N., Su, C.: An empirical study on the effect of individual factors on knowledge sharing by knowledge type. J. Glob. Bus. Technol. **3**(2), 1 (2007)

The Impact of Corporate Social Responsibility on the Firm's Financial Performance: The Case of Publicly Listed Companies in the United Arab Emirates

Sahar Mansour[1], Ali Mahdi Owayid[2](✉), and Muhieddine Ramadan[3]

[1] The British University in Dubai, Dubai, United Arab Emirates
smansour2000@gmail.com
[2] Al Buraimi University College, Al Buraimi, Oman
admin@buc.edu.om
[3] American University in Dubai, Dubai, United Arab Emirates
mzr@emi.ae

Abstract. The relationship between corporate social responsibility and the firm's financial performance is important topic for researchers. The primary aim of the research is to develop a new conceptual framework that will address the underlying relationships between corporate social responsibility and Financial Performance in the publicly listed companies in the UAE. The interaction effects between the components of corporate social responsibility and firm financial performance will be studies by responding to the main research question by way of a triangulation of research data within a mixed-method research paradigm that integrates both quantitative and qualitative methodology. The result of analysis should dictate whether or not a "Knowledge Gap" exists between the academic (theoretical) and the practical (applied) domains.

Keywords: Corporate social responsibility · Financial performance · Companies in the United Arab emirates

1 Introduction

In financial literature it is traditionally understood that the fundamental goal of corporations is to enlarge profits. Yet, over the most recent years it is widely discussed whether firms, which have officially enshrined goals relating to issue in the owners' benefit, also have a bigger responsibility reaching to the community as a whole. These responsibilities include the employee, environment, customer, supplier, Corporate Governance as well as wellbeing of the stakeholders. Usually there exist a lot of major contradictions between environmental or social goals on one side, and commercial objectives on the other. This takes us to the pertinent question – "What is the relationship between corporate social responsibility and Firm's Financial Performance for companies listed in the UAE stock exchanges?" In spite of the vague relationship among corporate social responsibility and a corporation's financial performance, in addition to the dispute of the real requirement for the socially responsible activities, it is

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 209–220, 2015.
DOI: 10.1007/978-3-319-21009-4_16

known that a lot of firms are engaged in many activities relate to the corporate responsibility. Such activities involves non-financial yearly reports, governance and environmental initiatives of the firms, rules and strategies, business codes and the process of dealing with or controlling things or people [1].

There are diverse considerations with regards to social responsibility by firms. Some consider social responsibility as a virtuous obligation that generate abundant benefits, while others believe that it plays a negative role in generating profitability to businesses. In other words, it is still an ongoing debate. The issue of corporate social responsibility "CSR" was perceptible during the 1930s and progressively so up to the 1960s. After that, Merrick Dodd at the Harvard Law School and Adolf Berle at Columbia Law School carried out a concentrated debate. Their argument focused on main question, which was "for whom are corporate managers trustees?" Dodd believed that the existence of corporations is imperative in providing societies with social services besides their objectives in generating optimal profits. Berle on the other hand criticized this and totally disagreed.

The intensity of the debate progressively slackened during the 1980s due to the rouse of hostile takeovers and gain after the dissolution of the Soviet Communism. This lead to an increased attention and focus on social responsibility, whereby social responsibility has been becomes a key issue not only for businesses but for theories and practices of law, management, economics and politics. Advocates of Corporate Social Responsibilities in today's world claim that CSR raises and encourages the ethical behavior of managers, which in turn positively impact the reputation of the corporation and its profitability [2].

2 Literature Review

The substantial increase in consumptions to encourage the corporate social responsibilities in the earlier periods of time encourage shareholders to find out a financial advantages from corporate social responsibilities initiatives, mainly taking into the accounts the financial goal of the firm as maximising the owner's wealth. However, empirical analyses of corporate social responsibility (CSR) and financial performance (FP) initiated before about thirty years ago and the findings of these analyses have been confused. Palmer [3] claimed that there are triple findings for the association among corporate social responsibility and financial performance: positive relationship, no relationship, and negative relationship.

The theoretical and empirical analyses proposed the first association: that there is a positive relationship between CSR and FP. Jones [4] claimed that by adopting the instrumental stakeholder theory, then the stakeholder satisfaction will impact the FP of the company. However, the instrumental stakeholder theory is combining two theories. First, the instrumental theory is a financial theory that expected the findings that will be revealed as a result of shareholders management. The next theory, the stakeholder theory, is a moral theory that suggests decision makers have an obligation to put stakeholders' requirements first for the purpose of increasing the value of the company, so that there is a positive relationship between the CSR and FP.

Studies carried out by Aupperle et al. [5] suggested the second option: that there is no association between CSR and FP. They used return on assets measurement in both short-run and long-run on a sample of 241 managers. Another study which is made by Ullmann [6] to prove that there is no relationship between CSR and FP. Ullmann [6] estimated that there are so many intervening factors between CSR and FP, and he also claimed that there are still various problems in measuring the intangible implications of CSR.

Lopez [7] was one of the most famous advocates about the negative relationship between CSR and FP, he has made a study for two years in 2002 to evaluate the financial performance of the firms by measuring the profit or loss of them, and he used the Dow Jones Index and the sample of 110 European companies. However, his study has a limitation as he analysed the short-term relationship between CSR and FP. Friedman [8] also claimed that CSR and FP have a negative association. He concluded that a “firm is an artificial thing” and so not be able to have a real responsibilities.

The contemplated research will include an exhaustive literature review with a concentration on current academic research in the field of corporate social responsibility and firm financial performance in order to cover the most important theories and seminal authors from the following perspectives: origins of corporate social responsibility research, development of its definition and classification, determinants of corporate social responsibility, development of the valuation models, development of studies examining the corporate social responsibility-performance relationship. Important and well-cited authors within the field will be included.

The perception of social responsibility has progressed through time in three movements. The first one is identified as the social responsibility movement, which considers that the responsibility of the firms merely focuses on the firm’s business obligation and on motivation. The second movement is known as the social responsiveness movement, which stresses on actions and activities conducted by firms in order to meet specific social obligations. It emphasizes activities that have expected and distinct outcomes that match or provide developments of policies relevant to the firm.

Finally, the third movement, which is the CSR practiced today and is well known as corporate social performance. It translates the effectives of corporate policies that are used to achieve social goals. It focuses on satisfying a variety of stakeholders such as employees, and the overall community. In other words, it holds the company responsibly towards the community as a whole besides its normal objective in maximizing profits [9].

A large body of literature review has attempted to identify determinants of corporate social responsibility. The significant determinants are rated variables related to community, Corporate Governance, customers, suppliers, employees, environment, business ethic, and controversies.

Earlier studies reveal that relationships between the corporate social responsibility variables and firm financial performance exist [10]. Referring to the earlier variables listed herein above, the following are these again with the drivers that define their ratings.

2.1 Community

This rating evaluates the firm's attention with respect to the community. Hence, for a firm to score positively, it must:

- Demonstrate an ability to actively disclose relevant activities through certain reports that are accessible by the public. Thus, building and maintaining a profound relationship with local communities.
- Devise behavioral principles and policies with regards the communities.
- Establish formal systems or departments in order to manage the relationships with the public.

2.2 Corporate Governance

Corporate governance is essential towards the protections of the rights of the stockholders, mainly those who possess minority shares. Usually firms gain extra points if the management of the firm is willing to strengthen its Corporate Governance policies and fully disclose their effectiveness. Hence, the relevant qualities of these may include the following:

- The total number of directors including independent and non executive directors as well.
- The frequency of board meetings
- The availability of audit committees and best practices
- The directors' compensation system
- The shareholders' voting rights in different classes
- Potential current disputes due to voting rights or compensation issues with regards to directors.

2.3 Customers

This rating is used to measure the attention of the firm towards its customers pre and post sales of a good or service. Presales imply the investigation of the needs and requirements of customers to identify the right product. However, post sales takes into account the level of customers' satisfaction. The following are used usually as a rating measure:

- Market research is periodically conducted and hence a relationship with the general public is already established.
- Warranties are of certified qualities.
- Acquired customers hold and established relationship with the firm.
- Products are guaranteed to be safe.
- Research to validate the level of satisfaction customers usually reveals.

2.4 Suppliers

Under this rating scheme, a company usually scores high, if it applies the following:

- The utility of formal procedures in selecting suppliers
- Visiting suppliers on periodic basis
- Fairly negotiating with suppliers
- Maintain an existing relationship with third world firms.

2.5 Employees

As far as this rating is concerned, it is mainly used to verify whether the company utilizes proper policies, rules, and managerial systems with regards to the characteristics to follow:

- Ensuring that the workplace is healthy and safe
- Demonstrating respect for employees
- Exercising freedom of relevant words and associations
- Active participation in unions
- The chance for employees to participate in making decisions and the chance to have sharing of profits
- Training and developing employees as much as needed.

2.6 Environment

The following are usually used to assess the attention of the firm with respect to the environment:

- The ability and willingness to apply the required procedures in protecting the environment, such as the reduction of water consumption and pollution, with an extra effort in recycling the raw materials.
- Managerial systems to be always certified
- Being complaint to existing laws
- Produce with minimum impact on the environment.

2.7 Business Ethic

With regards to the business ethic rating, it mainly deals with tools that are used by the company in an intention to preclude corruption, dishonesty and illegal actions and to foster righteous practices.

2.8 Controversies

This indicator reveals existing cases related to the environmental practices by a firm especially when quality of a product is in jeopardy. For example, any production of products that promotes pollution will affect this rating from a negative perspective.

On the other hand, there are many variables that could be used to measure the firm financial performance, such as market value/book value ratio, or the market value added model [10]. In this study, the firm's operating income will be the variable to the measure the firm's financial performance.

3 The Research Question

"What are the characteristics of a new conceptual model that assists in explaining the relationships between corporate social responsibility and firm financial performance for companies listed in the UAE stock exchanges?". In order to adequately respond to the main research question the following null hypotheses have been developed to provide quantitative benchmarks and data points for further consideration within the qualitative analysis of the subject matter. The null hypotheses below are chosen to study the interaction effects between the components of corporate social responsibility and firm financial performance.

Null Hypotheses (Table 1):

Table 1. Null hypotheses

$H1_0$	There is no significant relationship between Community and FP.
$H1_1$	Community positively affects firm FP for publicly listed companies in the UAE
$H2_0$	There is no significant relationship between Corporate Governance and FP
$H2_1$	Corporate Governance positively affects FP for publicly listed companies in the UAE
$H3_0$	There is no statistically significant relationship between customers and FP
$H3_1$	Customers positively affects FP for publicly listed companies in the UAE
$H4_0$	There is no statistically significant relationship between suppliers and FP
$H4_1$	Suppliers positively affects FP for publicly listed companies in the UAE
$H5_0$	There is no statistically significant relationship between Employees and FP
$H5_1$	Employees positively affects FP for publicly listed companies in the UAE
$H6_0$	There is no statistically significant relationship between environment and FP
$H6_1$	Environment positively affects FP for publicly listed companies in the UAE
$H7_0$	There is no statistically significant relationship between business ethic and FP
$H7_1$	Business ethic positively affects FP for publicly listed companies in the UAE
$H8_0$	There is no statistically significant relationship between controversies and FP
$H8_1$	Controversies positively affect financial performance for publicly listed companies in the UAE

4 Related Works

According to Page [11], a company achieves better financial performance once shareholders have more power than management. Another body of literature review supports this claim focusing on the ownership structure, the institutional shareholders in particular, being transient and caring more on short-term earnings, thus putting more and more pressure on managers to provide high levels of earnings. The chosen independent variables would be extremely relevant to the research. For instance, it has been concluded that there is a positive correlation between the ownership structure and the firm's financial performance [12]. The board does not take the role of running the day-to-day operations of the firm. Alternatively, the board is involved in major corporate decisions and simply delegates other responsibilities to the corporate officers. There are many vital duties that must be carried-out by the board periodically. One of its extremely important goals is to ensure that the firm's financial condition is accurately reported to its shareholders whilst conducting activities that enhance the profitability of the firm and the firm's share price [13]. From that perspective, it can be deduced that the size of the board has a positive correlation with the firm's financial performance. Empirical evidences of the relationship between corporate social responsibility and firm performance with respect to the earlier literature are mixed [14]. To the best knowledge of the researcher, gained through preliminary investigation of the academic field, it is believed that research faces a shortage of studies that measure the effect of corporate social responsibility on the financial performance of firms. It is believed that the scope of this research does not appear to have been completed in the United Arab Emirates yet, which provides opportunity to contribute original knowledge to the domain of Corporate Governance and financial performance research.

5 Contribution to Existing Knowledge

Quality research employs a sound research design to deal with important problems. Original research must add to the existing literature body by contributing to theory, knowledge, methodology, and/or practice [15]. The research goals should be clearly linked to the theoretical and physical context of the research and to its intended contribution [16]. This section aims at the context and background for the research and the extent to which the proposed research builds on what is known. The focus on corporate social responsibility has witnessed further attention not only within academia but laterally in the outside world as well. In the other words, corporate social responsibility and its importance is considered interdisciplinary. Researchers from different areas of interest such as Economics, Finance, Accounting, Management, and Law have been conducting various researches aiming to explore its relevance and importance.

On the other hands, evaluation and analysis of CSR accountability needs an action of who is assumed to be accountable to whom and this is the fundamental component of principal-agent models. In principal-agent theory, some actor called an agent carried out some activities on behalf of another actor called a principal. However, the principal can make actions that affected the agent's incentive to take any of its various actions. This procedure of structuring the agent's incentives is the main point of principal-agent theory. The decisions carried out by the principal that structure the incentive of the agent to take various actions constitutes the principal-agent theory [17].

5.1 Theoretical Contribution

Empirical evidence does not, consistently support measurement and disclosure theories, and the results found appear to be contradictory [1, 18]. To the knowledge of the researcher gained through a preliminary investigation of the literature it is believed that detailed research focused on testing the interactive effects of corporate social responsibility elements on financial performance in the publicly listed companies in the UAE has yet to be completed. Moreover, a primary aim of the research is to develop a new conceptual framework that will address the underlying relationships between corporate social responsibility and firm financial performance.

Figure 1 shows the expected contribution of the contemplated research, where the constructs: Community, Corporate Governance, Customers, Suppliers, Employees,

▪ Dependent Variable: Financial performance measured by firm's Operating Income ▪ Independent Variables: Community, Corporate Governance, Customers, Suppliers, Employees, Environment, Business Ethic and Controversies
Source: Author

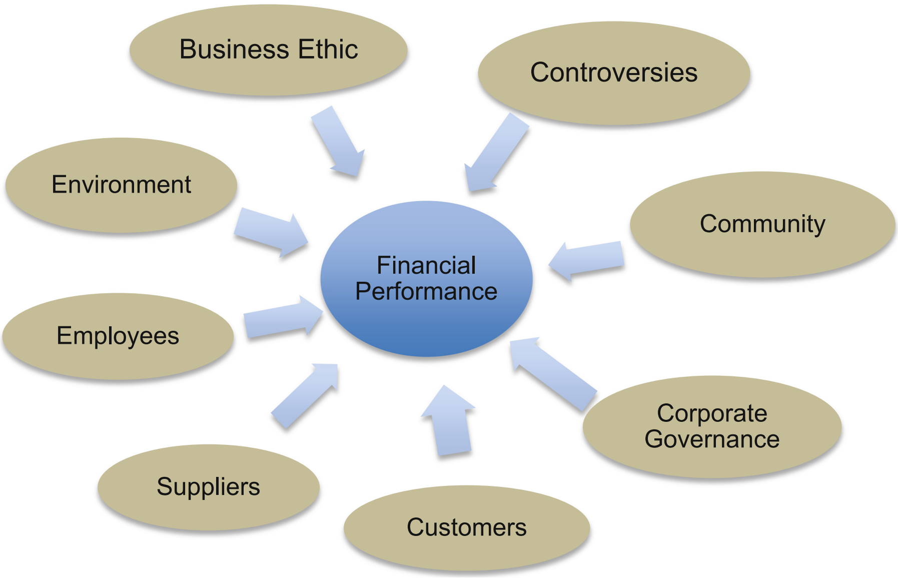


Fig. 1. The impact of corporate social responsibility on firm financial performance

Environment, Business Ethic, and Controversies are integrated into a model, which elucidates the impact of corporate social responsibility on Financial Performance.

6 Research Methodology

The contemplated research is designed as an empirical study of the relationships between corporate social responsibility and firm financial performance for publicly listed companies in the United Arab Emirates. The aim of the present research is to respond to the main research question by way of a triangulation of research data within a mixed-method research paradigm that integrates both quantitative and qualitative methodology, the research will be conducted through the following steps:

Step One: Desk Research – In-Depth Literature Review Part 1. Working from the preliminary literature review carried out for the purposes herein an in-depth review of the seminal authors within the domain of corporate social responsibility will be the first step completed with the research process. Part one of the In-Depth Literature Review will provide a solid foundation on which the Content Analysis of Step Two may be considered and placed into an overall scientific context both from the perspective of academic as well as professional knowledge. The literature research process in general has been broken into two parts for illustrative purposes. In reality, the literature review is an on-going iterative process that will continue in parallel as the content analysis of Step Two and the Semi-Structured Interview process of Step Three are carried out and finalized. The In-depth literature review will be returned to in a more formal manner in Step 4 in order to provide more academic detail in the exploration of the "Gap" analysis stage.

Step Two: Desk Research - Content Analysis. Based on the foundational readings of the Literature Research – Part 1 mentioned above, a quantitative analysis considering a series of null hypotheses based on data obtained from annual reports, supporting commercial documents and other data sources will be considered in order to develop an empirical understanding of the relationship of Corporate Governance to firm financial performance, i.e.: independent versus dependent variables. Data analysis will be concluded using statistical packages such as SPSS.

Step Three A - Field Research - Semi-Structured Interviews. With the benefit of the Literature Research completed in Step One and the results of the content analysis from Step Two the field research, which will take the form of a qualitative semi-structured interview process with key strategic stakeholders within the industry. The development of the interview questions will be informed by results of the Literature Review and Content Analysis steps. Stakeholders to be considered will be industry participants at a Vice-President position and above with a minimum of 5 years of industry experience within the UAE stock exchange listed companies. At present, there are approximately 130 UAE listed companies. Sample size is to include a minimum of fifteen (15) unique participants from varying companies with a possibility to extend to a maximum of twenty-five (25) unique participants, resources permitting. Unique participant is defined as one respondent from each unique company who will be permitted to participate in the interview process in order to ensure data integrity and safeguard from error due to duplication of responses. Interviews will be held in person at a location amenable to the subject and are expected to be thirty (30) minutes in length. Telephone interviews will be used in the case that physical interviewing is impossible due to resource or time constraints. Interviews will be tape recorded unless objected to by the participant in which case manual notes will be taken. The results of the semi-structured interviews will be analyzed using a statistical package such as NVIVO.

Step Three B – Field Research – Follow-Up Interviews. If both time and resources as well as participant engagement allow follow-up interviews of a more specific and narrow view informed by the first round of interviews, content analysis and literature review to date will be concluded with a smaller sub-set of respondents from the First Round sample, but not to exceed ten (10) individuals. It is anticipated that the follow-up interviews having the benefit and insight of the First Round Interviews, previous content analysis and literature review to date should provide an even deeper understanding of the relationship between Corporate Governance and firm performance by uncovering additional personally held beliefs and subtle understandings of the critical factors considered with respect to Corporate Governance in the commercial day-to-day environment of the selected participants.

Step Four – Desk Research – In-Depth Literature Review – Part 2. Once the second round of interviews have been completed and a more granular appreciation of the issues faced by the sample participants is uncovered through the statistical analysis of both the quantitative (SPSS) and qualitative data (NVIVO) a return to the Literature Research will be completed to further refine the scope and consideration of the existing knowledge within the academic field. This step is necessary and critical in order to ensure that the scope of the research is as delineated as possible and that the most critical, important and recent literature is considered going forward.

Step Five and Step Six – Triangulation of the Data & Gap Analysis. With the benefit of the both the desk and field research being completed in the form of content analysis, two-stage participant interviews and exhaustive literature research a triangulation of the data will be considered and analyzed in order to determine whether or not the existing academic knowledge is congruent with the practical application of the field on a day-to-day commercial basis. The result of this analysis should dictate whether or not a “Knowledge Gap” exists between the academic (theoretical) and the practical (applied) domains.

Step Seven: Development of New Model. Building on the Gap Analysis in Step Six and determination of whether agreement is shown between the practical application and the theoretical analysis of the seminal literature a thorough analysis of the existing frameworks within the academic domain will be made.

This analysis will inform whether or not the existing frameworks sufficiently address the requirement for practical application existing within the industry and whether or not they may be further improved or modified. The same analysis if showing significant disagreement between the theoretical models and the practical application found within industry will dictate the development of a new model or framework as an alternative

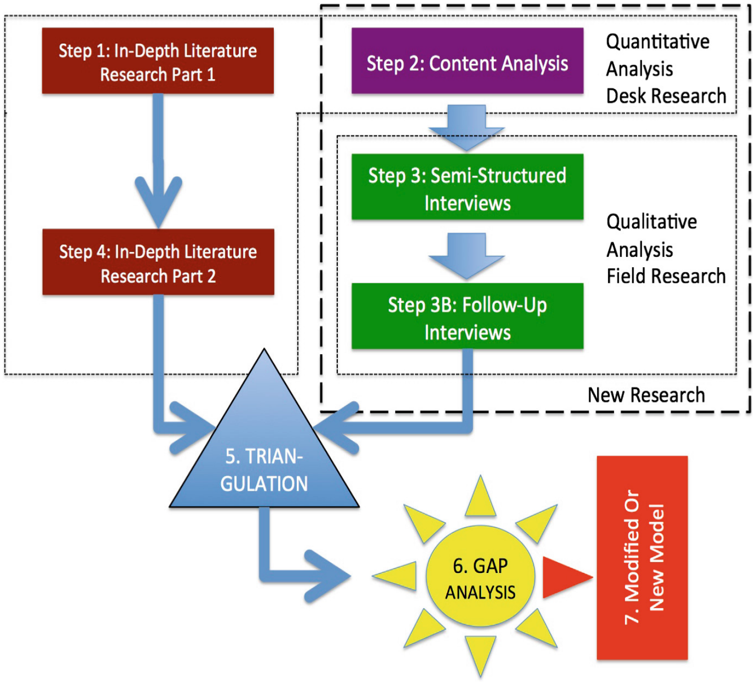


Fig. 2. Research process flow

methodology for corporate social responsibility with the intent of bridging this "Gap" between the theory and practice. This final step in the research is considered to be the contribution of original knowledge taking the form of the development of a uniquely modified or significantly new framework or model. The above Research Steps are illustrated in Fig. 2.

7 Conclusion

Thus far, the preliminary literature review suggests that harmony on the definition of corporate social responsibility has proven to be illusive and that the industry can be qualified as pre-paradigmatic. In addition, there is still a mixed perception between various researchers regarding the effect of corporate social responsibility on a company's financial performance. The magnitude and direction in which corporate social responsibility can affect the company's performance depends on the framework and industry type. Thus, the importance of the contemplated research is to draw this relationship in a conceptual framework and to try to bridge the gap in knowledge that appears to be present.

References

1. Tronsgaard, T.: The relationship between corporate social performance and corporate financial performance: an event study of firms listed on OSEBX. Doctoral dissertation, BI Norwegian Business School (2014)
2. Iwu-Egwuonwu, R.C.: Does corporate social responsibility (CSR) impact on firm performance? A literature evidence, social science research network (SSRN) (2010)
3. Palmer, H.J.: Corporate social responsibility and financial performance: does it pay to be good?. http://scholarship.claremont.edu/cmc_theses/529. Accessed 2 April 2015
4. Jones, T.M.: Instrumental stakeholder theory: a synthesis of ethics and economics. Acad. Manag. Rev. **20**, 405–432 (1995)
5. Aupperle, K.E., Carroll, A.B., Hatfield, J.D.: An empirical examination of the relationship between corporate social responsibility and profitability. Acad. Manag. J. **28**, 446–463 (1985)
6. Ullmann, A.: Data in search of a theory: a critical examination of the relationships among social performance, social disclosure, and economic performance of US firms. Acad. Manag. Rev. **10**(3), 540–557 (1985)
7. Lopez, M., Garcia, A., Rodriguez, L.: Sustainable development and corporate performance: a study based on the Dow Jones sustainability index. J. Bus. Ethics **75**, 285–300 (2007)
8. Friedman, M.: The social responsibility of business is to increase its profits. http://reason.com/archives/2005/10/01/rethinking-the-social-responsi. Accessed 6 April 2015
9. Orlitzky, M.: Corporate social performance: developing effective strategies. Research Brief RB004. Australian Graduate School of Management (2000)
10. Venanzu, D., Fidanza, B.: Corporate social responsibility and value creation: determinants and mutual relationships in s sample of European listed firms. Social Science Research Network (SSRN) (2006)
11. Page, J.P.: Corporate Governance and Value Creation. The Research Foundation if CFA Institute, VA (2005)
12. Bowen, R.M., Rajopal, S., Venkatachalam, M.: Accounting discretion, corporate governance and firm performance. Social science research network (2004)
13. Kim, K.A., Nofsinger, J.R., Mohr, D.J.: Corporate Governance, 3rd edn. Pearson, New York (2010)
14. Baker, K., Anderson, R.: Corporate Governance. A Synthesis of Theory, Research, and Practice. Wiley, New York (2010)
15. Wellington, J., Bathmaker, C., Mcculloch, G., Sikes, P.: Succeeding with Your Doctorate. Sage Publications, London (2009)
16. Andriessen, D.G.: Making Sense of Intellectual Capital. Elsevier Butterworth-Heinemann, Amsterdam (2004)
17. Bolton, P., Dewatripont, M.: Contract Theory. MIT Press, Cambridge (2004)
18. Phusavat, K., Comepa, N., Sitko-Lutek, A., Ooi, K.B.: Interrelationships between intellectual capital and performance: empirical examination. Ind. Manag. Data Syst. **111**, 810–829 (2011)

Identification of the Main Problems in the Management of Innovation Processes and the Draft of Appropriate Recommendations

Viliam Lendel(✉), Štefan Hittmár, and Eva Siantová

Faculty of Management Science and Informatics, University of Žilina, Univerzitná 1, 010 26 Žilina, Slovak Republic
{viliam.lendel, stefan.hittmar, eva.siantova}@fri.uniza.sk

Abstract. Management of innovation processes in company is the field of innovation management which is still not enough researched and applied in practice. Managers in companies often do not know about modern techniques and design tools for creating innovation processes and also as about the possibility of their effective usage for management and effective usage in decision-making conditions. The purpose of this paper is following a detailed analysis of literature and realized research to create a model of innovation processes management in the company. There are also identified the main fields of potential problems, which should be discussed by managers to achieve effective functioning. The solution of the questions researched within the article needs to use several methods depending on the character of particular parts of the solution.

Keywords: Innovation · Innovation process · Innovative ideas · Management · Company

1 Introduction

The issue of management of innovation processes is currently very actual. Innovations are an important tool for increasing competitiveness of companies. Companies do not develop their innovation activities on the basis of "impressions" or "intuition", but on the basis of knowledge obtained from the opinion survey of customers, employees and partners. They collect the necessary information and innovative ideas, reveal innovative opportunities and make decisions about the need to innovate. The aim of the companies is to realize their full potential for innovation. However, in order to be successful, it is necessary to effectively manage these activities and to be able to quickly and flexibly respond to developments in the market. Right there is a place for the identification of major problems and draft of appropriate recommendations to ensure the effective use of market opportunities through innovation.

The main purpose of this paper is to gain new knowledge in the field of innovation management with a focus on the identification of the main problems arising from the

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 221–232, 2015.
DOI: 10.1007/978-3-319-21009-4_17

management of innovation processes and to point out to the possibility of proposing appropriate recommendations to minimize them. The identification of the major problems can significantly contribute to improving the management of innovation processes in Slovak companies. The recommendations should serve as a valuable tool for managers to a successful course of innovation processes in the company.

Solving of the examined questions in the paper requires the use of several methods, depending on the character of the various parts of the solution. For gathering and collection of information the analysis (the analysis of current and historical data on the issue), questionnaire method and semi-structured interviews (data collection in empirical research) were used. In the processing of information the quantitative evaluation method (statistical techniques) and the comparison method (when comparing data obtained by empirical research and analysis of data from secondary sources) were used. In order to solve the problem the methods of induction, deduction, synthesis (in the identification of the main problems arising in the management of innovation processes and the formulation of appropriate recommendations), abstraction and modelling were used.

2 The Current State of Dealing with the Issue

Innovation process can be considered on the base of scientific literature analysis as an organized and controlled sequence of activities where inputs, in the form of innovation ideas, are transformed into outputs, in the form of innovations. It is a process of recognizing customer needs and innovative opportunities, generating innovative ideas and their elaboration, work with information and knowledge regarding innovation, realization of innovative activities and ensuring successful extension of innovation among customers [7].

Since new customer needs are created on the base of extension and usage of a new product, an innovation process can be understood as a repeating process [13]. Furthermore the number of innovation processes is not limited. Innovation process has also a built-in mechanism of learning, i.e. a mechanism evaluation of incurred failures and deviation in all phases of the innovation process [12].

The mechanism is supported by feedback in all phases of the process [12]. This mechanism is also supported by its openness which allows the company to adopt innovative ideas, necessary licenses for research and development from the external environment [5]. In the case of non-utilization of own innovation solutions, it should offer these solution to other companies, in the base of license.

In the scientific literature [1, 3, 5, 10] it is possible to find a number of theories trying to explain to managers companies how to create innovations and which factors influence the result of this process. These theories look on the innovation process from different points of view while the effort to create the complex view of innovation activities in a company can be considered the integrating element. In many cases, however, this effort leads to excessive complexity and ultimately to confusion, which often affects the decision-making of manager of the company.

Several authors [1, 4, 9] point to the fact that the issue of innovation processes run diverse research that is still disunited and inconsistent. Creation of a unified view to the innovation process is significantly impeded by the following factors:

- innovation management is based on a number of scientific disciplines and it cooperates with them,
- the objects of research are various types of industries and markets, which have their own specific effects and signs,
- the object of interest of theorists and practitioners are various types and forms of innovation, such as technological, organizational, products, process, and so on.

These are the main disadvantages of current work. On the base of the detailed analysis of the various approaches to the innovation process management in the company by various authors as well as the evaluation of the practical approaches can proceed to the systematization of lessons learned. This is a summary of the main benefits of different innovation process models that provide resources for creation of a comprehensive model for the innovation process management (Table 1).

Skokan [11] highlights the growing importance of regional innovation systems, which enable easier sharing of tacit knowledge and increase the capacity for localized learning. A factor to promote innovative activity is capacity of identifying opportunities that allow for changes and new business inside and outside the company through alliances and strategies supported on the use of ICTs, so that the organization becomes pioneer in its action field [2].

3 Results of the Empirical Research

Between October 2012 and January 2014 we conducted a research, whose primary goal was to gather and interpret information about the level of use of innovation processes management in the environment of Slovak enterprises. In total, 321 managers of small, medium and large enterprises from companies active in Slovak republic took part in the research. Calculated recommended sample size was 384 respondents. The survey covered 321 respondents. Following the conversion, the actual sampling error was at the level of 5.46%.

The survey focused on identifying situations in the various phases of innovation process. The first field of interest was represented by innovative ideas which may accrue from a variety of sources. In terms of Slovak companies, customers and their identified needs are the most frequently used source of innovative ideas (in 228 companies). The most frequently used sources of innovative ideas also include employees of the company (in 194 companies), analysis of competitive products and services (in 187 companies), Internet (in 175 companies), exhibitions, conferences, trade shows (in 166 companies), journals and publications (in 159 companies). As a significant source of innovative ideas can also be considered own research activities (115 companies) and the research of the partners of the company (in 98 companies).

Managers of Slovak companies indicate customer needs analysis as a source of the most successful innovative ideas, respectively ideas with the greatest potential. Customer focus should also be implemented in the management of innovation processes.

Table 1. Benefits of different models of innovation processes

Author(s)	Emphasis	Fortification
Imai et al. [6]	Functional integration	Integrated innovation process
Rothwell [10]	Electronic data processing	Information assurance of innovation process
Dvořák et al. [3]	Feedback and creative activity	Cross connection of innovation process
Vlček [13], Mol and Birkinshaw [9]	Needs dissatisfaction with status "quo"	Needs/dissatisfaction as the main launcher for innovation process
Bernstein and Singh [1]	Manifestation of management (control element)	Organization of the innovation process for site management
Tidd et al. [12]	Timing and learning	Innovation process base on a continual improvement
Skokan [11]	Systematic approach Role of environment	Innovation system (national, regional) Local innovation process
Chesbrough [5]	Level of openness	Bounded (closed) Unbounded (open)

As resources for successful innovative ideas the respondents considered also their own research activities and their own employees. In this case it is necessary to finance and support their own research activities, create a favourable environment for their workforce and care for their educational development.

The importance of customers confirmed the fact that only 257 companies (80.06%) create innovative ideas based on the input from customers and their requirements. Another stimulus for generating innovative ideas is finding a specific problem (186 companies, 57.94%). This means that these companies prefer the opposite strategy, it means they do not apply a proactive approach, companies create ideas only in the case of arising problem. Other less used incentives include business needs (81 companies, 25.2%) and new technologies (72 companies, 22.43%). As a positive can be considered that only 9.03% of companies generate the innovative ideas at random and only 5.92% of Slovak companies do not generate them at all.

Companies in Slovakia do not use an information system to work with innovative ideas (recording, sorting, distribution…). This was confirmed to 61.28% of managers surveyed. The information ensuring of innovation process has only 28.96% of respondents, the most used information system is a system provided by the company Salesforce.com (16.72% companies). Other companies use their own IT solution. In the survey 9.7% of the respondents were not able to express their opinion about the use of the information system to work with innovative ideas.

Managers of Slovak companies identified as the main criteria for deciding on further elaboration of innovative ideas into innovative opportunities availability of funds (involved in decision-making to 25.81%). Other important decision criteria are the reality of demand, technological options, available knowledge in the issue and the availability of human resources. In terms of priorities for decision-making are time and the physical space in the development phase less involved.

The most used methods in the deployment phase of innovation are the techniques of development of the creativity (42.99%). The quite frequently used methods are conceptual methodological tools (24.61%), forecasting methods (19.63%) and pragmatic methodological tools (19.31%). On a small scale are also used techniques of knowledge management (14.64%) and innovative graphs (7.79%).

As major problems hindering the effective management of innovation processes by the managers are considered: the lack of the necessary financial resources to ensure the innovation process (210 companies), distrust of the company managers to the possible outcomes arising from a lack of innovation (187 companies) and missing information ensuring the innovation process (168 companies).

4 Management of Innovation Processes in the Company

Management of innovation processes in the company is a real problem, which are managers of the company faced. Number of factors participates in its provision, from theoretical concepts, through model solutions to practical applications. The biggest or the most significant problem can be seen in an absence of unified or complex and at the same time transparent management innovation processes model in the company. It can be said, based on the results from realized research, that on the present many companies try to manage their innovation activities and processes intuitively. However, companies must often solve various problems caused by their unpreparedness to manage innovation processes. A requirement to create a comprehensive and transparent innovation processes model, which would be a significant aid for managers, resulted from interviews with several representatives of companies within realized research. The benefits from this model would appreciate not only managers, but also customers, because their communication with company related to preparation and assurance of products and services by fulfilling customers' requirements and needs. This can lead to the elimination of a number of problems both from companies and its customers.

A management innovation processes model in the company (Fig. 1) was designed on the basis of views of various authors dealing with the issue of creating a model for managing innovation processes in the company and also after careful analysis of mentioned approaches to the development of innovative processes. Solution is based on a management of innovation portfolio model according to Hamel [4]. His definition of innovation has two approaches. First it represents competences that need to be built, and second it is a process that needs to be implemented in the company. Therefore is model made up from two main parts, specifically: innovation process and innovation competences.

Innovation process is a sequence of activities aimed at creation and implementation of innovation. It includes activities related to generating innovative ideas, their evaluation, creation of innovation and ensuring its spreading among customers. A transparent model of the innovation process was developed to better understand its singe phases. It is based on the basic model by Zaušková and Loučanová [14]. However, it is enhanced by identified key elements. Specifically, it is about adding the first phase of

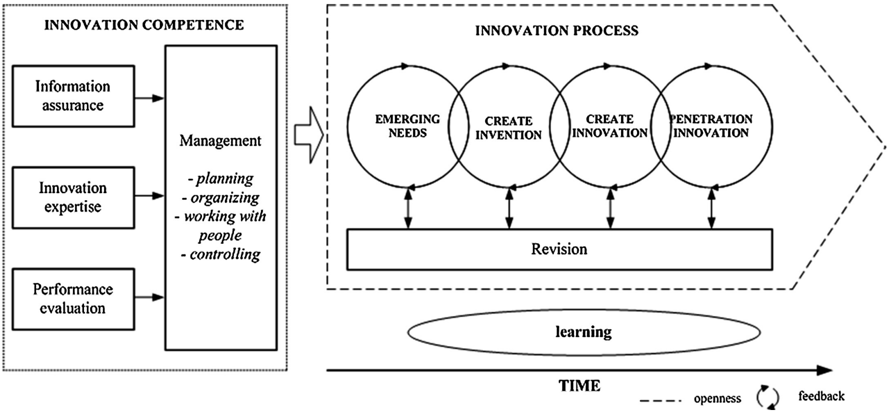


Fig. 1. Management innovation processes model in the company [7]

the innovation process aimed at creating the need, or dissatisfaction with the status "quo". Need, or dissatisfaction is perceived as the main trigger of the innovation process. This fact was highlighted by authors Vlček [13], Mol and Birkinshaw [9]. Furthermore, in the innovation process is regarded the degree of openness, of which importance noticed Chesbrough [5]. It is a boundless innovation process that allows flexible work with innovative ideas, which come to the company both from the internal and from external environment. Company can offer unused innovative ideas to other businesses by licensing, and vice versa, if necessary, it may acquire innovative ideas from external environment. Next element is feedback, which is highlighted in the innovation process. It allows to overcome problems accrued in the various stages of the innovation process. This is an approach proposed by Dvořák [3] in a form of chain links in the innovation process. Innovation process takes into account also the element of learning, which should contribute to a continuous improvement of the process. Tidd [12] considers learning as a critical point in the management of innovation. Company can learn through the procedure in the innovation process and thus improve its management methods.

According to Hammel [4] innovation competence are understood as a tools which allows company to use in innovation process management innovation tools, supported communication and information technology, management processes and appropriately to establish criteria for measuring the success of the individual phases of innovation projects:

- Innovation expertise,
- Information security,
- Management (planning, organizing, leading people, control),
- Evaluation of the level of innovation processes management and innovation performance.

On the base of the proper usage of various management functions (planning, organizing, leading people and controlling) the innovation process should be effectively managed. The communication is important role in this process. An essential foundation for the successful management of innovation processes in the company can be considered an effective system of information security innovation processes. Another innovation competence is innovation expertise. It is necessary that the company management is supports the creative human potential. This means that employees how to fully use their skills and capabilities to ensure the innovation process.

In order the innovation processes should be managed effectively it is necessary to report some results in fields which affect to the management of innovation processes. Every company has a different level of management of innovation processes. Therefore, it is necessary to identify in the first step the current level of innovation process management in the company, then discover weaknesses and make recommendations for their improvement. It is necessary to build up an appropriate methodology for evaluating the innovation process management in the company.

5 Identification of the Main Problems and the Proposal of Suitable Recommendations

The management of innovation processes in the company is for managers a challenging task. Causes of failure in this process can be multiple and may have a different character. For example, there may be a lack of innovation expertise, failure to secure information flow in the company, lack of education and motivation of employees and so on. Reasons of failure in the management of innovation processes are affected several actors. Firstly, they are managers, in the case of lack of support to innovative activities in the company, employees of the company, in the case of passive participation in the innovation process and customers, in the case of indifference in providing an added.

The management of innovation processes cannot be carried out only intuitively based on the development of the situation. This is a complex process with a numbers of aspects: the state of innovation potential, built IT infrastructure, staff evaluation system, organization and so on.

The task of these recommendations is to help reduce the risk of identified problems. They are designed to help the managers in the management of innovation processes, as well as prevention of the occurrence of problem situations. These can arise in the following areas:

- Problems in the information security of innovation processes,
- Problems in ensuring the innovation expertise,
- Problems in the application of management elements in the innovation process,
- Problems in measuring of innovation performance,
- Problems in the innovation process and the process itself.

Problems in Ensuring the Information Security in Innovation Processes. As a problem arising in this field can be considered the inefficient information flows in the company. This is the reason of misunderstandings and employees cannot realize their innovative tasks entirely.

It is recommended to the managers to ensure the efficient work with information related to innovation in the company. This means that it is necessary to ensure the access to information to all interested parts in the innovation process, and it is necessary to collect all valuable information in one database and apply the principles of effective work with information.

Another problem is the lack of information systems to work with innovative ideas, opportunities and innovations. In many cases there is no evidence of implemented innovations and innovative ideas generated in the company. This often means that the potential innovative opportunities are left unused.

It is recommended to the managers to make a comprehensive record of innovative ideas and innovations in the company. Every innovative idea in the company have to be recorded and assessed. Same importance has the record of the currently unusable innovative ideas which have potential value for the future.

Problems in Ensuring the Innovation Expertise. One of the problems in this area are the personal characteristics of managers and employees of the company focused on emphasizing their ego, what is seen in the subjective view of the solutions to the problem and ignoring acquired facts and reality. The result is incorrect decisions that adversely affect the conduct of the innovation process.

It is recommended to managers to evolve personal characteristics oriented to empathy, teamwork and also to logical thinking and self-criticism.

Insufficient implementation of knowledge management belongs to the common problems in the field of innovation expertise. On the one hand it is reflected as a lack of knowledge of the employees, on the other hand, much of the knowledge created in the innovation process is forgotten or lost. The result is unnecessary, repeated creation of new knowledge which is already created in the innovation process.

In the first case it can be recommended to the managers to ensure the management training courses designed to supplement the necessary knowledge to employees. In the latter case it is necessary to provide recording and archiving of acquired knowledge in the innovation process through the implementation of appropriate IT solutions.

In the area of ensuring the innovation expertise can be seen as a problem an inadequate usage of creative thinking in the development of ideas to solve the problem. The result is low number of ideas that do not allow to take a decision to resolve the problem.

It can be recommended to the managers to develop human creative potential in the company. It is necessary to create conditions for application of creative thinking selecting appropriate exercises to develop creative skills and abilities of employees. It would also be appropriate to create innovative teams, including representatives of the young and old. Innovative team should include representatives of several departments of the company (production, sales, logistics, trade, service).

Problems in the Application of Management Elements in the Innovation Process. Problem is the lack of a coherent methodology for the management of innovation in the company. The managers can often exchange the innovation process by using of a simple type of creative brainstorming techniques. In many cases, they are in time and work stress and they do not pay attention to the support of innovation and integrating innovations into long-term strategic plans of the company.

The managers may be encourage to pay more attention to innovative activities of the company, turn them into long-term business objectives and incorporate them into innovative business strategy. It is also recommended to attending educational activities objected to the managing of innovation processes.

Another problem is the lack of development of the innovation program. In many cases, managers do not have sufficient information about available resources and means when they plan innovative activities. The result is the increased probability of failure of implementation of the innovation project.

It is recommended to carry out a detailed analysis of the current state of innovation potential and application of methods and techniques of project management.

Another problem is the absence of remuneration for innovative ideas and appropriate motivation program. The result is the passivity of employees who are not motivated to bring new innovative ideas and engage in innovative task solutions beyond their tasks and responsibilities. It is recommended to the managers to establish a fair system of remuneration for innovative ideas. It is necessary to develop appropriate motivation program, which involves employees in innovation activities of the company. Employees will be informed of the expected changes and motivation program will encourage open communication within the company.

Failure of the management of innovation processes can be caused by unsuitable organizational structure, which does not allow open communication between the stakeholders and does not support new innovations. It is recommended to the managers to rethink the current organizational structure, creation and implementation of flexible organizational structure that will have the ability to respond to changes in business environment and allow fast exchange of information, organization of meetings and activity of innovative teams.

Problems in the Measurement of Innovation Performance. Acommon problem in this area is non-evaluation of effects and benefits of implemented innovations. The result is that the company has no feedback on the adoption of innovations by the customers, cannot measure their performance and take measures leading to continuous improvement of the management of the innovation process.

It is recommended to the managers to create an evaluation system focusing on the diagnosis of the results and contributions created and on the market launched new products. Based on the results it is also recommended to formulate measures to improve the management of the innovation process.

Problems in the Innovation Process and the Process Itself. Acommon problem is, that the first and the second phase of the innovation process is carried out insufficiently. In many cases it can be observed the development of the first innovative idea that pops up without detailed analysis. The result is the frequent changes and increased costs of implemented innovative projects. It is recommended to the managers to carry out a detailed analysis of all the innovative ideas and suggestions from internal and external environment. It is necessary to pay attention to the first two phases of the innovation process because they affect its success. Only the identification of valuable innovative ideas can bring successful innovation.

Another problem is the application of an inappropriate model of the innovation process, which reflected to a lack of a clear definition of the problem, the lack of coordination of activities, communication and cooperation within the staff of the various departments, but also of the stakeholders. The result is the unsuccessful management of innovative projects. It is recommended to the managers to be careful in choosing the model of the innovation process. It is necessary to know the company conditions, the level of knowledge of employees, the information security, the set of business processes and so on. There is a place for the usage of feedback in each phase of the innovation process, which may indicate deficiencies caused by improper setup.

The main problems include the lack of a mechanism for learning, which evaluates errors and variation, documented knowledge, rules and principles. Without its usage the continuous improvement of the management of the innovation process cannot be ensured. It is recommended to the managers to actively use the knowledge, closely monitor the progress of the innovation process and learn from the results of the different phases.

6 Management on Innovation Processes in Selected Company

The selected company for the presentation of practical management of innovation processes is KROS, a.s., which deals with the development and distribution of economical, constructional and expert software since 1995. The main identified problem is the inadequate information security of innovation processes, particularly in respect of effective work with innovative ideas. These ideas are obtained from various sources (customers, employees) and they are recorded in different information systems, which complicates their usage, because of potentially good idea which might go unnoticed or duplication of ideas. Many information are not recorded in the IS and they are available only to certain employees and stored in their minds, which prevents their proliferation and usage. Subsequently the rationalization of innovation processes was designed:

- all innovative ideas, obtained from customers and employees will be recorded in one common information system,
- innovative ideas will be evaluated in the system by more employees and customers.

In product development it is important to understand that freedom stimulates creativity, and therefore employees should be respected and their creativity should be encouraged. Subsequently, the follow principles of creating innovations were recommended to the company:

- integration of multiple perspectives, it means integration of sales employees, development, consultants and customers and together create products that will meet the needs of customers. To unite these people in one team across several departments.
- Highlighting the importance of the human factor in the management of teams, where the leader should be the element that starts innovative thinking within each team. For this it is advisable to use SCRUM.
- Emphasis on added value for the customer and understanding the vision of the company in which is the satisfaction of customer in the first place.

The mentioned recommendations aim to continue in the trend of increasing the number of customers (currently over 74,000) and to maintain the leading position in the Slovak market.

7 Conclusion

Effective management of innovation processes should identify weaknesses (gaps) and take measures to eliminate them [8]. It should also be capable of delivering the necessary information related to innovation to responsible persons. A company should be prepared in the management of innovation processes to certain risks that may arise and cause failure of the realized innovation projects. If the company can identify these risks as soon as possible and prepare for them, it will significantly increase the success of management of innovation processes in the company.

In solving the defined problem were identified key weaknesses (gaps) and main difficulties of the innovation process management in terms of Slovak companies (empirical research realized by authors). Valued results of this study are also formulated recommendations how to correctly manage innovation processes in a company.

As main conclusions can be mentioned: Management in the company is important in the field of innovation processes management. A key assumption for the successful realization of the innovation processes is the existence of a supportive environment for creation of innovations. For managers can be recommended to implement and use the system of self-evaluation of innovation processes in the company.

Acknowledgements. This paper was partially supported by the Slovak scientific grant VEGA 1/0363/14 Innovation management – processes, strategy and performance.

References

1. Bernstein, B., Singh, P.J.: An integrated innovation process model based on practices of Australian biotechnology firms. Technovation **26**, 561–572 (2006)
2. Piraquive, F.N.D., García, V.H.M., Crespo, R.G., Liberona, D.: Knowledge management, innovation and efficiency of service enterprises through ICTs appropriation and usage. In: Uden, L., Oshee, D.F., Ting, I-Hsien, Liberona, Dario (eds.) KMO 2014. LNBIP, vol. 185, pp. 300–310. Springer, Heidelberg (2014)
3. Dvořák, J., et al.: Management inovací. Vysoká škola manažerské informatiky a ekonomiky, Praha (2006). ISBN 80-86847-18-7
4. Hamel, G.: Leading the Revolution, 1st edn. Plume, New York (2002). ISBN 0-452-28324-8
5. Chesbrough, H.W.: Open Innovation: The New Imperative for Creating and Profiting from Technology Hardcover, p. 272. Harvard Business Review Press, Watertown (2003). ISBN 978-1578518371
6. Imai, K., Nonaka, I., Fakeuchi, H.: Managing the new product development. In: Clark, K., Hayes, F. (eds.) The Uneasy Alliance. Harvard Business School Press, Boston (1985)
7. Lendel, V.: Management of innovation processes in the company. Habilitation thesis, University of Zilina: Faculty of Management Science and Informatics (2014)

8. Lendel, V., Varmus, M.: Evaluation of the innovative business performance. Procedia Soc. Behav. Sci. **129**, 504–511 (2014)
9. Mol, M.J., Birkinshaw, J.: The sources of management innovation: when firms introduce new management practices. J. Bus. Res. **62**(12), 1269–1280 (2009)
10. Rothwell, R.: Industrial innovation: success, strategy, trends. In: Dodgson, M., Rothwel, R. (eds.) The Handbook of Industrial Innovation. Edward Elgar, Aldershot (1994). ISBN 9781858984452
11. Skokan, K.: Konkurenceschopnost, inovace aklastry vregionálním rozvoji. Repronis, Ostrava (2004). ISBN 80-7329-059-6
12. Tidd, J., Bessant, J., Pavitt, K.: Řízení inovací, p. 549. Computer Press, Brno (2007). ISBN 978-80-251-1466-7
13. Vlček, R.: Strategie hodnotových inovací. Professional Publishing, Příbram (2011). ISBN 978-80-7431-048-5
14. Zaušková, A., Loučanová, E.: Inovačný manažment, p. 91. Technická univerzita, Zvolen (2008). ISBN 978-80-228-1953-4

Innovations for Competitiveness

Knowledge Management Model for Project Management

Flor Nancy Díaz Piraquive[1(✉)], Víctor Hugo Medina García[2], and Rubén González Crespo[3]

[1] Catholic University of Colombia, Bogotá, Colombia
fndiaz@ucatolica.edu.co
[2] District University "Francisco José de Caldas", Bogotá, Colombia
vmedina@udistrital.edu.co
[3] Pontificial University of Salamanca, Madrid, Spain
ruben.gonzalez@upsam.net

Abstract. Currently, organizations are aimed at developing, organizing, and using knowledge to create successful projects. Similarly, these organizations consider knowledge an important asset that collaborates in complying with the company's strategic objectives and, hence, it is necessary to manage, control and maintain it. This article presents some knowledge management models that provide relevant factors for project management to suggest a model that avails of their strengths and best practices, to obtain satisfactory results based on experiences and learning identified.

Keywords: Organization · Management models · Knowledge management · Strategic objectives · Project · Project management

1 Introduction

Knowledge is interpreted as the virtue that can help to understand and interact with the performance environment, it is developed through "own or transmitted" experiences and helps to solve problems that emerge daily within a competition environment. "Knowledge is the whole set of cognitions and abilities with which individuals tend to solve problems; it comprises theory and practice, daily rules, as well as the instructions for action. Knowledge is based on data and information, but – unlike these – it is always tied to people; it is a unique part of individuals and represents their beliefs about causal elections" [1].

The prior definition shows that knowledge is part of all people, but it is often not sufficiently or adequately exploited, which leads to repeating the same experiences or the same processes without any benefit and, which is worse, making the same mistakes. Knowledge management has been mentioned since the 1990s to group several knowledge areas and manage them optimally, exploiting them to the maximum to give added value to the different "tangible and intangible" assets people have. Knowledge management applied to *project management* seeks, once the project's path has been set, to support it to obtain good results within reasonable times and in optimal manner,

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 235–247, 2015.
DOI: 10.1007/978-3-319-21009-4_18

based on the experience acquired during planning, execution, follow up, and control and close of previous projects.

This article proposes a knowledge management model that enhances project management, availing of relevant and important factors analyzed in some knowledge management models, which have been recognized for their results, highlighting the strengths and best practices applied to obtain satisfactory achievements based on experiences and learning identified within organizations.

The structure of the document took into account the theoretical foundation, formulation of the problem and objectives, methodology, data collection, analysis an result, conclusions and future research and references.

2 Theoretical Foundation

Knowledge is a source of resources that in themselves do not produce any result. With this premise, it is possible to see the importance of managing this resource and its relevance in large enterprises or companies, which are always searching for best practices, both internal and external, to benefit their performance. Currently, the experience has demonstrated that the use of knowledge management applied in project management becomes necessary, not only to control its execution, but also – upon defining the direction in which the project is to be guided – for it to be developed optimally, avoiding failures or risks that can be mitigated by analyzing past errors, which were socialized because they were present in similar projects and which were documented as best practices in the “memoirs of a project”, which are managed by the Project Management Office (PMO).

Likewise, it is evident that Project Managers in their quest for being leaders, document the project’s results, along with the experiences and knowledge acquired, to generate new knowledge that team members can appropriate and use ahead in similar projects or developed simultaneously. In this sense, authors North Klaus and Roque Rivas discuss on how to manage knowledge when used by a company or intelligent organization, generating the following questions, which are worth bearing in mind for its good use: Are stakeholders sensitized, especially business directors, workers, and investors, with respect to the meaning of the knowledge resource for business success? Which are the strategic objectives that should be supported as a priority through the mobilization of knowledge? Which are the necessary abilities today and which in the future to ensure sustainable and competitive capacity? How can we identify the barriers that hinder the exchange and development of knowledge within the company? How should we configure and develop the company for it to now and in the future to increase its competitive capacity [2].

Currently, knowledge management is used by many organizations as competitive strategy to achieve their positioning, as well as maintaining themselves and being recognized in the market as companies that meet their objectives and generate quality and innovative goods or services. In recent years, knowledge management has become a valuable and important resource; thereby, it must be guided, managed, controlled, shared, and transferred [3]. Knowledge management is divided into two groups: measurement of intellectual capital and knowledge management. Intellectual capital can be defined as: “Intellectual capital is the appropriation of knowledge, applied experience, organizational technology, relations with clients and professional skills, which give a company competitive advantage in the market" [4]. "It is the set of intangible assets of an organization, which although not reflected in traditional financial statements, currently generate value or has the potential to generate it in the future" [5]. Stated otherwise, it is the knowledge people have or can generate to create or give added value to activities or processes.

2.1 Knowledge Management Models

This research socializes the following five knowledge management models considered the most relevant to the respect:

2.1.1 Spiral Model (Nonaka and Takeuchi)

These authors propose a model for the process of creating knowledge, describing it as a four-phase process in spiral shape (Universidad Nacional Abierta y a Distancia [6]. In this model, Nonaka and Takeuchi talk about tacit knowledge and of explicit knowledge, separating each into two big pillars. This model seeks to verify if good use is being made of the processes and if the personnel is the most suitable for the organization, permitting measurement of the intangible resources to manage them.

2.1.2 Knowledge Management Model by KPMG Consulting

This model is based on two fundamental aspects for knowledge management; the first is the organization's learning capacity and the second factor is the results expected from said learning. For said learning to occur, it is necessary for all the members to integrate and show desire for learning. The results expected are: permanent change, more competent action, people development, and construction of the environment. All founded upon the culture, commitment, learning capacity, and on the interrelation among people, teams, and the organization [7]. It identifies six basic elements the organization must direct to learning and which are not independent; on the contrary, they interact amongst themselves: strategy, organizational structure, leadership, personnel management, information and communication systems, and the organizational culture. The factors that compose the capacity to learn are:

- Firm and consistent commitment from the whole company, clearly led by the company's upper management because this process requires involvement from all parties.
- Behaviors and learning mechanisms at all levels, that is, propitiate an environment with mechanisms for the creation, catchment, storage, transmission, and use of knowledge [8].
- Development of the company's infrastructures in favor of strengthening operations and the behaviors of individuals to maintain change and permanent learning.

2.1.3 Model of Knowledge Management by KMAT

The purpose of this model is to achieve for every leader to develop the strategies and definitions of the business objectives. So that it allows them to direct and control the use of its resources in function of knowledge. It is an organizational administration model centered on knowledge, which identifies, organizes, applies, shares, and permits the creation of flexible structures that facilitate the transference of knowledge; maintaining an organizational culture that increases trust among its members and the exchange of ideas; promoting integration of units, team work, and tolerance with mistakes, favoring learning, innovation, and flexibility against changes. The model proposes four facilitators (leadership, culture, technology, and measurement), which favor the process of managing organizational knowledge [9].

2.1.4 Model of Knowledge Management, Arthur Andersen

Anderson [9] speaks of knowledge management as the integration of the individual and the organization, referring to the importance of human contribution, but that it will not be availed of correctly if the organization does not implement infrastructure and processes and promotes a culture that permits the analysis of the individual's knowledge [10]. According to [11], this model has the following definitions:

- The generation of value with clients is given by the organization's ability to optimize the flow of information.
- Stemming from the individual, there is the responsibility to share and make knowledge explicit for the benefit of the organization.
- From the organization, there is the responsibility of generating a suitable climate and provide tools in terms of infrastructure, culture, and technology that give way to the knowledge process.

Additionally, the company must focus on constantly implementing new technologies that take and maximize knowledge, as well as the implantation of methodologies, memories of knowledge or repositories. If any of these two aspects fails, the chain will break and the model will not yield the results expected [12].

2.1.5 Model of the Intellectual Capital Value (Hubert Saint and Onge)

Implemented for the first time in the Canadian Imperial Bank, this model studies what relation there is among intellectual capital, the organization's learning, and its measurement. This model is totally internal and global; it does not measure the relationship between the organization and its individuals or the organization and the external "client" [13].

This model is composed of three elements: human capital, structural capital, and client capital. "Human capital determines the structural and it, in turn, influences upon the client capital, as well as on the financial aspect, but indirectly" [14].

3 Formulation of the Problem and Objectives

Currently, many of the projects do not achieve the totality of the objectives proposed due to factors like time, estimation of resources, or poor definition in its reach. This generates a negative impact upon the organization, which must decide to assume the costs this creates in time, money, productivity and even reputation. Often, the decision is to liquidate the project, classifying it as a total failure, with the consequential loss of opportunity of the market or client served.

One of the recurrent causes that generate a project's failure is that of not using methodologies, models, or tools that help with knowledge management in projects, making project managers to wonder on: what practices, methodologies, or models should be used to guarantee the project's success without falling into the trap of looking for and acquiring solutions that do not fit the business needs. If an effective knowledge management model existed, it would be quite useful to respond forcefully to said question; however, the lack of a model of knowledge management becomes an error that increases the problem in achieving the project's objectives.

According to the Project Management Institute (PMI), currently organizations are not implementing models, techniques, or tools that support knowledge management to accomplish effectiveness in project management. Statistics reveal a drop in the use of models from 2010 to 2012. This behavior is shown in Table 1 [15].

Also, based on the Chaos report by the Standish Group, the entity in charge of gathering information on the failures presented in technology project management, which shows in the report published in 2013 the decline in the results of project management since 1995 to 2012.

As indicated by the figures, a slight improvement of successful projects has been noted since 2009 to 2012, but – even so – the results are disappointing, given that 61 % of the projects in 2012 failed. It cannot be understood why results in the execution of project still remain low – despite the availability of sources, techniques, tools, and models that facilitate the generation and transference of knowledge, like those analyzed in this article, as well as the existence of knowledge guides like PMBOK, PRICE2 along with the ease of application of monitoring tools like MS Project (Table 2).

According to the figures of success factors of a project, proposed by the Chaos Report, presented in Table 3, it is important to recognize that project success is represented in a high percentage by the specialization of human talent or project team,

Table 1. Statistics on the use of models, tools, and techniques in project management. Source: [15].

% of the organizations that currently have	2010	2011	2012
Formation in tools and techniques of project management	65%	62%	59%
Processes to develop skills in project management	52%	50%	45%
Processes to mature existing project management practices	51%	47%	44%

Table 2. Results in project management 1995–2012. Source: [16].

% of results in project management	1995	2009	2012
Successful	15%	30%	39%
Failed	55%	45%	18%
Cancelled	30%	25%	43%

Table 3. Success factors. Source: [16].

Success factors	Points
Support from upper management during execution	20
Participation of users	15
Optimization of resources	15
Experienced team	13
Experience in project management	12
Agility in processes	10
Clarity in business objectives	6
Emotional maturity	5
Execution	3
Tools and infrastructure	1

(tacit and explicit knowledge) and experience in project management. Due to this, to diminish the amount of failed projects, it is necessary to implement a knowledge management model for project management that contributes in its success, through sharing and transferring knowledge gathered in the best practices and by learning from mistakes, which capitalized and documented by PMOs. The project manager has a big challenge when trying to manage projects in such a manner to guarantee increased participation of projects successfully executed.

All the models that talk about knowledge management have as sole base "knowledge" that is catalogued as a resource of value for companies or organizations. In spite of this, its structures are completely different, depending on the author's point of view, which shows that no method is infallible. Hence, the general objective is to define a model of knowledge management applicable to project management, bearing in mind the existing theory and referenced in these types of models. However, some specific objectives are proposed, like:

- Conduct a comparative analysis of the knowledge management models described in the theoretical foundation to identify their strengths and weaknesses.
- Apply knowledge acquired with the description of the models and combine it with the experience acquired in project management.
- Propose a knowledge management model for organizations that involve project management, taking the most relevant aspects from the models studied and applying them in the construction of the new model.

4 Methodology

Throughout the development of this research, the documentary management methodology was used. The documentary management methodology is part of the scientific research process, which seeks to solve or explain a particular type of problem.

> "*Documentary research seeks to request information to recognize, gather ideas or suggestions, locate and define problems that permit fine tuning the methodology to conduct more adequate and complex research, without arriving at definite or general conclusions*" [17].

That is, the documentary management methodology seeks to gather, select, analyze, and present results based on the use of documents. Among the advantages of using this methodology are the ease of finding relevant information and the capacity to generate evidence with experiences from other authors. The following are the stages of the documentary management methodology used [17]: election and definition of the theme and determination of subthemes, selection of information, organization of la information, report draft, and structure of the report for its presentation.

Additionally, for this research, the process model was used to solve the proposed information for the Research of the Catholic University of Colombia, which allows locate, select, collect and synthesize information. This model consists of five phases: (1) Definition of the problem. The problem focused on constructing a model KM to PM, where sharing facilitators, knowledge transfer ownership and is used. (2) The sources were: virtual journals indexed among others, Eric, Dyalnet, SpringerLink, Proquest, Ebsco, e-book, e-book. (3) Development of search strategies. The criteria used to perform searches relate to the fact that articles, books, documents, etc. ISBN or ISSN counted with. (4) Obtaining documents. The documents that met the criteria were downloaded considering the chosen source and date of consultation. (5) Assessing, evaluating information and publishing. This last phase corresponds to the results and discussions are presented below [3].

5 Data Collection, Analysis, and Result

Upon analyzing the characteristics and definitions given for each model of knowledge management, a comparative table was prepared of it to highlight strengths and weaknesses (Table 4). Based on this, we expect to take the best practices to propose a new knowledge management model for project management.

Taking the aforementioned as reference framework, a knowledge management model is proposed, which can be applied in any company or organization that develops or executes different projects to solve its own needs or those of its clients. The model proposed is composed of several components, the first is the interrelation of knowledge, the second is the execution and management of projects, the third component is the integration process among the previous components, and the fourth component refers to the management elements. The following describes each of the components and will show the model is completed with each of these.

The *interrelation of knowledge* is illustrated as a cyclical chain. Knowledge is unique to each individual and it is given by their experiences, that is, it is tacit, but for it to be applicable, it is necessary to share it and bring it to the explicit part to conceptualize and contextualize it; hence, knowledge is disseminated and appropriated by a group that has shared knowledge in common. When knowledge is shared with the group, it is possible for other individuals to complement it, debate it, or accept it. This confrontation will lead to refining knowledge.

When the group consolidates and assimilates knowledge, a mechanism should exist to place refined knowledge at the organization's disposition. When knowledge is part of the organization, if adequate strategic guidelines exist, correct knowledge management is guaranteed, converting knowledge into a useful asset (Fig. 1).

Table 4. Strengths and weaknesses of knowledge management models for project management. Source: The authors

Model	Strengths	Weaknesses
KPMG	Learning capacity and obtain results from said learning	If members of a project do not integrate and do not show interest in learning, it is not possible to achieve results
	Propitiates settings for the generation, maintenance and use of knowledge	If there is no firm commitment from the company, the risk exists of not meeting the objective. This model proposes an interaction with all the elements
KMAT	Controls and guides resources in function of knowledge	Does not have knowledge measurement indicators
	Facilitates transference of knowledge	Does not implement tools for trasference of knowledge
	Promotes exchange of ideas among team members	
	Promotes innovation	
Arthur Andersen	Keeps in mind the contribution of the knowledge of people with the organization or project	If tools like infrastructure, culture, or processes are not provided, knowledge will not be availed of.
Hubert Saint and Onge	Has most-suitable personnel	Does not focus on the learning capacity. Knowledge is explicit for the benefit of the organization.
	Creates knowledge	Does not measure the relationship between individuals and the organization
	Bears in mind human, structural, and client capital	
Nonaka and Takeuchi	Propose tacit and explicit knowledge	Centered more on learning tan on acquiring suitable personnel
	Aimed at the creation of knowledge	
	Proposes growing spirals of knowledge	
	Has feedback and continuous improvement	

Fig. 1. Interrelation of knowledge. Source: the authors

This implies that knowledge is at the disposition of all the individuals who make up the organization, so that any individual who enters the organization, or is part of it, may obtain, update or generate new knowledge that expands the chain or keeps it updated and operational.

With the interrelation described, transference and Exchange of knowledge is guaranteed for the management model to be useful and effective for the organization, with the model being part of an improvement process to the extent that it is nourished. The second component of the model is the *execution and project management* (Fig. 2); in an organization different projects are managed to cover the needs of the internal or external clients. Within the project, knowledge is generated and consumed; knowledge that flows and may vary through each of the stages. Within project management the definition is formulated of the methodology through which projects are carried out, defining the flow of information in the project.

Independent of the methodology established for each project, it is indispensable to have feedback based on the implementation or execution of the projects. This knowledge management model expects the feedback to take place throughout the whole project; it is considered that during the whole execution potential use of knowledge is taking place, which is why it is valid to feed the lessons learnt at any moment of the project and once the project has finished verify and confirm said learning.

For this model, the methodology must include formalization mechanisms that make the connection of the knowledge generated or updated within the projects and knowledge as a useful asset of the organization, resulting from the first component, that is, it must seek commitment from the individuals who are part of the project to include their knowledge. This means the third component is precisely the *integration process* between the interrelation of knowledge and the management and execution of projects (Fig. 3). This process is not independent from the project or the organization; on the contrary, it is transversal to both, but it is a fundamental part of the model, its purpose is to place knowledge in the indicated moment to propitiate good practices, optimize, and increase the efficiency of the execution of projects to obtain better results.

This integration process is exactly the articulating part that drives the organization to establish a repository of lessons learnt and better methodologies and practices. It is suggested that for the creation of the integration process, the organization should have a strategic governance body of knowledge management of projects (PMO), which must

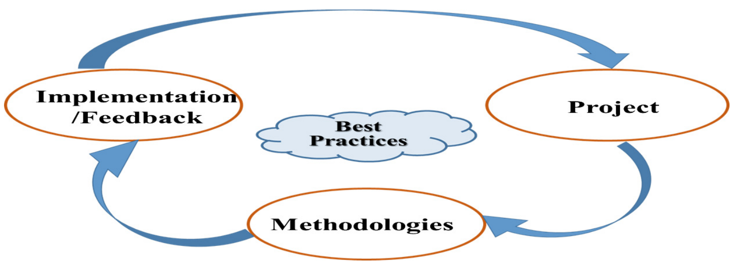


Fig. 2. Component of projects. Source: the authors.

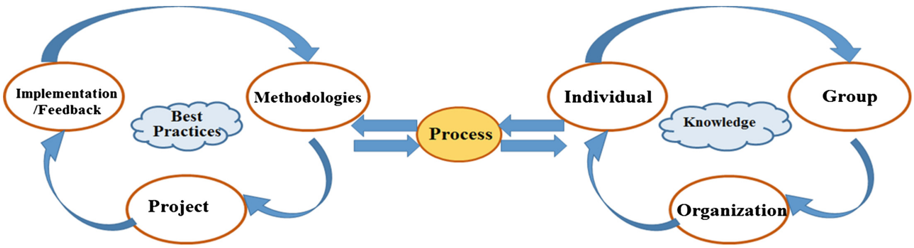


Fig. 3. Integration process of components. Source: the authors

guarantee the publication of knowledge among those involved in each project. This body must be in charge of compiling knowledge established through lessons learnt generated during each of the stages of the project and put together the corresponding documentation to feed the knowledge base. Likewise, with the availability of a specific area responsible for knowledge management, it would be an advantage that would facilitate the optimization of the methodology used to develop.

Another relevant aspect that should be contemplated during the integration process is the definition of indicators, although only one of the models analyzed includes the use of indicators. For the management model under development, it is considered that the use of said indicators makes it measurable and manageable, given that based on them it is possible to establish essential actions for the model's continuous improvement. Indicators or mechanisms should be implemented to measure, for example, the degree of training of the resources, the level of consultation of a project's lessons learnt, the quality and effectiveness of the documentation. These indicators will be established and revised by the PMO.

To the same extent, the model under development contemplates that it is not enough to obtain information, rather, it is also necessary to manage said information, so that it is of easy access for consultation by the organization. It should also allow controlled updating; all this is accomplished through managing the PMO information repository. Additionally, as mentioned by the KPMG model and by most of the models, it is necessary to have managerial support, given that the whole team's motivation to participate largely depends on them. Besides, they are in charge of propitiating the good use of the model to, then, obtain from it some solutions against situations that could have emerged previously in other projects. This support from upper management is part of the fourth and last component of the model, which refers to *elements of management* and with the knowledge management model proposed remains completely articulated.

Several of the models mention the requirement to provide the necessary infrastructure to support the model of knowledge management and this task is the responsibility of upper management that shapes the elements of management. Likewise, the organizational culture that leverages the documentation processes and the knowledge sharing practices are part of these elements (Fig. 4).

The organization's strategic guidelines, besides guiding and directing the course of the organization, are key for this model as elements of management that aid in defining the object of knowledge the model will have [18].

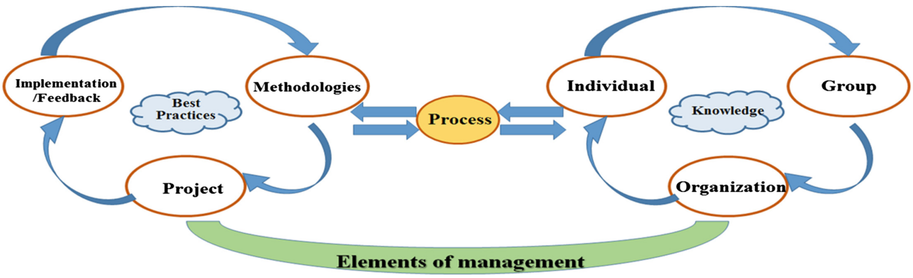


Fig. 4. Knowledge management model proposed. Source: the authors

6 Conclusions and Future Research

Knowledge is part of each person; to the extent that it is shared and applied in organizations, it gains greater value and becomes a useful asset or an intangible resource, which - thereby - must be managed efficiently in favor of the good results of companies.

It is important to recognize the different knowledge management models to use them correctly, depending on the type of project; it is even possible to take the best practices from each model to adapt a specific model to meet the needs of the company or organization in which it will be applied.

It is common in all models to integrate parties interested in the project that provide knowledge to it, given that every team has success or failure experiences.

Independent of the technological solution or the infrastructure used to implement any model, it is important to keep in mind what is required besides the organizational culture of learning and of sharing knowledge, as well as the establishment of clear and convincing processes for those involved to participate as expected in the application of the model.

It is important to consider this as an effective knowledge management model for project management, by applying positive experiences and avoiding the same mistakes made before with similar projects. This will help the project to be a success case and fit the business needs.

In every organization it is of vital importance to share and transfer knowledge in the area of project management for the organization to be more competitive and to meet its objectives at strategic levels. Additionally, knowledge will be enhanced and will gain greater value as an asset in a company.

Knowledge management is a key factor for the good development of projects, given that the experiences of the members from each of the company's lines of business provide for the projects to comply with each of their phases within the times agreed.

It is undisputed that the strong support of senior management is essential for achieving the project objectives, and to involve knowledge management processes in project management, begins with a change management process, which impacts the change in organizational culture.

Implementation of the PMO within an organization that develops projects is a big support to formalize and standardize a model of knowledge management, to guarantee management of a repository of lessons learnt and best practices and improve performance in different projects.

The articulation of KM processes with those of PM is achieved with the strong support of the PMO. It with input activators as incidents presented, errors drawn, unsolved problems, misused formats, poorly executed procedures and inexperience of human talent, among others, through KM enablers so much technological and learning as management systems, wikis, online messaging systems, coaching, mentoring, stories, workshops, teamwork, among others, will achieve results shown in best practices, lessons learned, error learning, standardization of forms, formats and procedures which would be for optimal performance of PM processes, managing to satisfy customer expectations according to the contracted scope, the programmed schedule, the allocated budget and to agreed quality.

Future work will focus on the further identification of obstacles to projects management and will explore the role of informal knowledge sharing in management innovation in projects.

References

1. Probst, G., Romhardt, S., Romhardt, K., Fernández, E.: Administre el conocimiento. Pearson Educación de México, Mexico (2001)
2. Klaus, N., Rivas, R.: Gestión del conocimiento. Una guía práctica hacia la empresa inteligente. Libros en rec (2008)
3. Díaz, F.N., Medina, V., González, R.: Global communication and innovation in service companies. In: Uden Lorna, F.O.H. (ed.) Knowledge Management in Organizations, pp. 300–310. Springer, Santiago (2014)
4. Edvinsson, L., Malone, M.: El capital intelectual. como identificar y calcular el valor de los recursos intangibles de su empresa. Gestión, Barcelona (1999)
5. Euroforum: Medición del capital intelectual. modelo Intelect. Euroforum Madrid (1988). http://gestiondelconocimiento.com/modelo_modelo_intelect.html
6. UNAD Universidad Nacional Abierta y a Distancia. http://datateca.unad.edu.co/contenidos/101110/EnLinea/leccin_1_modelo_creacion_del_conocimiento_en_las_organizaciones_nonaka__takeuchi.html (2010). Accessed 03 Sept 2014
7. Mertins, K., Heisig, P., Vorbeck, J.: Knowledge Management Concepts and best practices. Springer, Berlin (2003)
8. Tejedor, B., y Aguirre, A.: Proyecto logos: investigación relative a la capacidad de aprender de las empresas españolas. Boletín de estudios económicos. vol. LIII, No 164, pp. 231–249 (1998)
9. Andersen, A.: El management en el siglo XXI: herramientas para los desafíos empresariales de la próxima década. Granica, Buenos Aires (1999)
10. Publicar: Gestión del conocimiento (2010). http://publicar.wikispaces.com/Gesti%C3%B3n+Conocimiento. Accessed 1 Sept 2014
11. Lopera Londoño, M., Quiroz Gil, N.: Caracterización de un modelo de gestión de conocimiento aplicable a las funciones universitarias de investigación y extensión. Caso Universidad CES, Tesis de Maestría en Dirección, Universidad CES, Universidad del Rosario, Medellín (2013)

12. Orozco, L.D., Polvo Hernandez, K.: Tesis Propuesta para el tratamiento contable - financiero del Capital Intelectual en México, U. de las Américas Puebla, Mexico (2003)
13. Sánchez F.: Modelo de empresa innovativa. Vol. 250(3), pp. 321–334 (1999)
14. Cardona, Y.C., Correa, Z.C., Hurtado, C.D.: Aprendizaje organizacional: una capacidad de los grupos de investigación en la universidad pública. U. del Valle **26**(44) (2010). http://cuadernosadm.univalle.edu.co/index.php/cuadernosadmin/article/view/625
15. Project Management Institute: PMI's Pulse of the Profession (2013)
16. The Standish Group: CHAOS MANIFESTO (2013)
17. Hernandez, J.L.: Metodología de la Investigación en Ciencias Sociales. de VI. Fases de la investigación documental (2012). http://metodologiaencienciassociales.blogspot.com/p/vi-fases-de-la-investigacion-documental.html. Accessed 08 Sept 2014
18. Medina, V.H., Torres Acosta, J.H.: Modelo de gestión del Conocimiento para proyectos de ingeniería. In: Fifth LACCEI International Latin American and Caribbean Conference for Engineering and Technology, 29 May 2007

Development of Knowledge Model for Construction Projects

Andrej Tibaut[1(✉)] and Denis Jakoša[2]

[1] Faculty of Civil Engineering, University of Maribor, Maribor, Slovenia
andrej.tibaut@um.si
[2] Faculty of Electrical Engineering and Computer Science,
University of Maribor, Maribor, Slovenia
denis.jakosa@student.um.si

Abstract. The growing need for interoperability and the use of BIM (Building Information Modeling) within the AEC domain (Architecture, Engineering and Construction) is leading to the increased need for efficient knowledge management. In order to digitally manage and understand large volumes of information contained in construction works projects, the paper suggests a knowledge engineering approach with a common knowledge model. The new knowledge model (ontology) for AEC domain is presented and applied as a case study in a real construction project. As a result, a more effective knowledge management throughout the whole lifecycle of AEC projects is expected.

Keywords: Knowledge management · BIM · Ontology · Construction project

1 Introduction

Companies that have survived the economic crisis of the construction industry share belief in the importance of knowledge management in construction projects. In consequence the belief results in requirements for a more effective course of a construction project such as

- requirements for a more comprehensive management of data flow in construction projects by introducing an integrated information model (BIM) and
- requirements for more efficient knowledge management in the lifecycle of construction object.

The first requirement relates to the current culmination of an understanding of management of information throughout the lifecycle of a construction project (planning, design, engineering, construction, operations and maintenance), where stakeholders (architect, MEP and HVAC design engineer, structural engineer, quantity surveyor, suppliers, construction manager, site foreman, site supervisor and facility manager) traditionally communicate their documentation (data models), at best, as cumulative layered 2D CAD (DWG) drawings (schemas) or in worst case as PDF documents. The drawings (Fig. 1) for floor plans contain only drawing primitives in form of points and lines (plane XY) without elevations. The drawings (Fig. 2) for section plans (plane XZ and YZ) contain elevations but only for predefined cross-sections of the construction

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 248–259, 2015.
DOI: 10.1007/978-3-319-21009-4_19

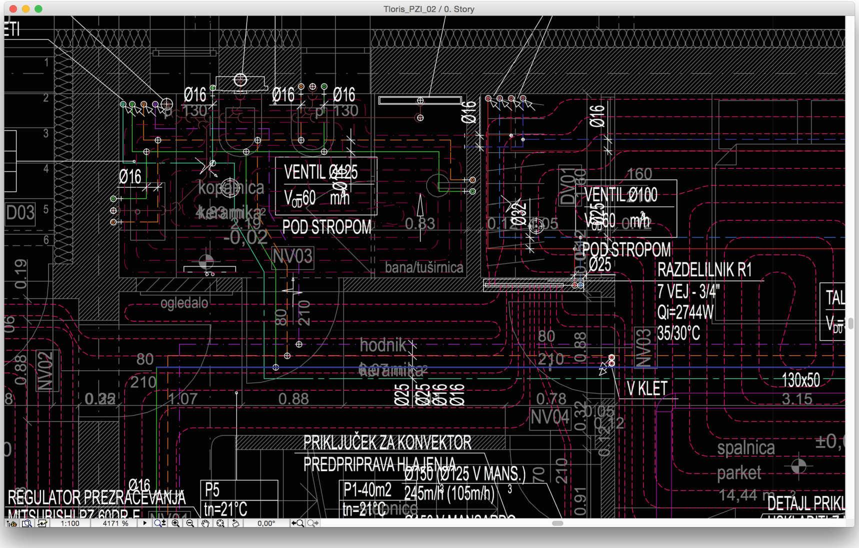


Fig. 1. Heating, mechanical and plumbing floor plan: drawing primitives with implicit semantic [13]

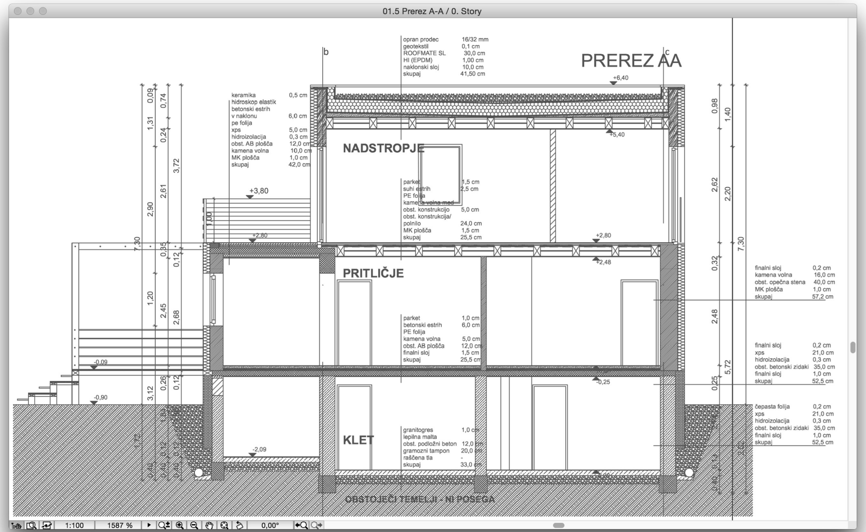


Fig. 2. Predefined section plan of a building: drawing primitives with implicit semantic [13]

object. Therefore multiple 2D CAD drawings must be consulted in order to perceive construction object in its 3D entirety. Finally, the semantics (description, physical dimensions, construction details, manufacturer, etc.) of the drawing primitives can be expressed with text-annotations only. The characteristics of the yet prevailing design and engineering work practices throughout the lifecycle of a construction object, which are relevant to the knowledge engineering, can be summarized as follows:

- Construction projects are fragmented among numerous stakeholders, which is also stated by other authors [1].
- Engineering designs in form of 2D CAD plans lack of an integrated information model.
- Information that is exchanged between stakeholders in the construction process does not necessarily have an understood meaning.
- Use of the weak knowledge assets results in lengthy procedures for preparation and coordination of the construction projects and, consequently, in unplanned costs due to limited digital interoperability between construction project stakeholders. Figure 3 shows unplanned cost as they are shared throughout the lifecycle of a construction object.

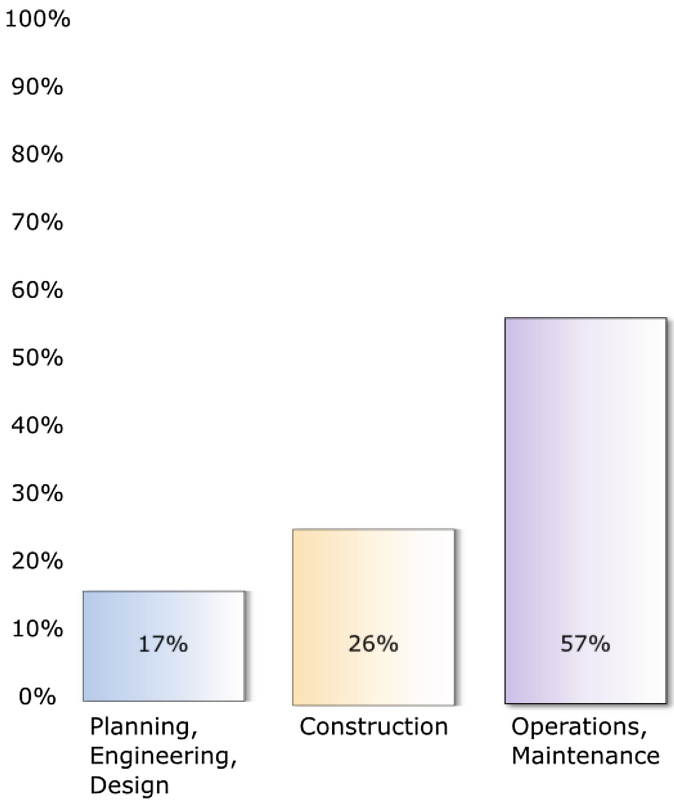


Fig. 3. Unplanned cost share because of inefficient digital interoperability in construction [7]

The aforementioned characteristics are believed to disadvantage the efficiency of construction industry in comparison to other engineering disciplines (i.e. automotive industry). Therefore a new work paradigm in construction industry, the Building Information Modeling - BIM [5, 11], is heavily influencing research and practice in construction industry. BIM could provide an ideal platform for integrated practice since it is capable of integrating and linking the project information to the data level [8].

The second requirement concerns the need for more efficient knowledge engineering in the lifecycle of construction object. Young civil engineers lack knowledge about the workflows in the lifecycle of a construction project but they demonstrate a good working knowledge of information and communication technology. In an effort to perform cost-effective and to meet economically reasonable deadlines for the construction project, all project stakeholders strive for the timely completion of works. In such circumstances experienced engineers (i.e. site foreman) practically do not have time to guide younger less experienced engineers (i.e. assistant site foreman) through the daily workflow activities (i.e. on construction site) in order for them to understand "why and how" [12].

To facilitate the knowledge exchange in construction project between experienced and less experienced engineers better personal knowledge management in connection to organisational knowledge management is required [4].

Inefficient digital interoperability leads to a non-existent or interrupted flow of information between stakeholders in the construction process (Fig. 4a). Interruption of the information flow results in the repetition (i.e. architect and structural engineer do not share integrated building model, instead they both create one for themselves) and high chance of errors.

According to research [6] 86 % of public works projects end up with cost overruns. Some unexpected findings of the research were that:

- Technically difficult projects were not more likely to exceed the budget than less difficult projects.
- Projects in which more people were directly and indirectly affected by the project turned out to be more susceptible to cost overruns.
- Project managers generally did not learn from similar projects attempted in the past.

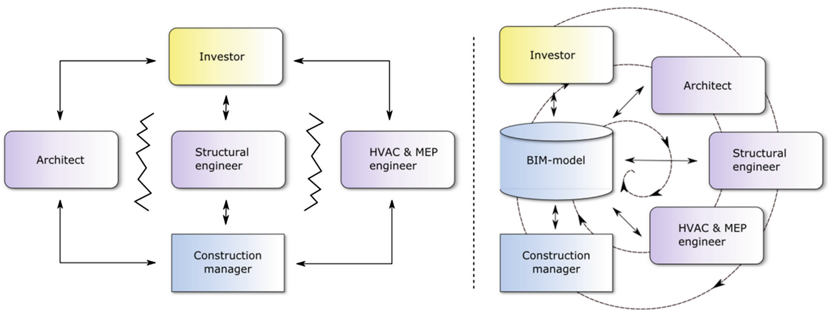


Fig. 4. a) Typical (interrupted) information flow in construction project. b) Information flow integrated with Building Information Model (designed by authors)

More efficient knowledge management in the construction processes is possible with common server based information model (BIM server, Fig. 4b). Such server stimulates an overall collaborative environment based on a common reference data model for AEC industry. Apart from existing commercial server based solutions, a viable and open source solution is BIMserver.org [2]. BIMserver.org's functionality includes selection, update, and deletion of (sub)models of a construction object (i.e. an architectural model, a model of mechanical installations) and it also conforms to the specification of Industry Foundation Classes (IFC) [9].

2 Case Study: Knowledge Model for a Construction Project

The growing need for interoperability and the use of BIM within the AEC domain is leading to the increased need for efficient knowledge engineering and management systems.

In order to control, manage and understand large volumes of information in the two current (2014, 2015) infrastructure construction projects "Underpass Grlava" and "Underpass Ljutomer", which are carried out within the project named "Electrification and reconstruction of the railway line Pragersko – Hodoš", we have looked for solutions in knowledge modelling and semantic technologies. On the Fig. 5 the construction site of the construction project "Underpass Grlava" is shown.

As part of our case study methodology we have decided to investigate common characteristics of the two construction projects. This can be best achieved with existing technical documentation (i.e. consultants' technical report). The technical documentation summarizes all important facts about the construction project.

For the application of semantic technologies we have decided because the two construction projects have the following characteristics:

- A large amount of construction data for "Underpass Grlava" and "Underpass Ljutomer" (technical report, detailed design plans, construction schedule plan) that we want to model syntactically and semantically;
- Technical reports for both projects contain rich, but non-uniform and hard-to-relate, structure and technical terminology, which are the results of the work of two different designers (i.e. use of the term "tact" and "lamella" for a scheduled zone on the construction site). We want to align the available data from the technical report at the level of the common concepts.

As a result, we expected a more effective management of knowledge on the project's construction site.

The project involved one structural design engineer and one draughtsmen working in the engineering consultancy office, a project manager, site foreman/manager, assistant site foreman, several labourers, specialists and sub-contractors working on-site.

At first we had to develop a knowledge model that defines the structure of the data using an open source ontology editor and framework for building intelligent systems Protégé.

Fig. 5. Construction site "Underpass Grlava": use case for application of knowledge model

While creating a knowledge model we relied on a standardized data dictionary initiated by BuildingSmart [3]. The result was a general knowledge model (ontology) that can be universally applicable in AEC projects.

3 Knowledge Model – Construction Ontology

For the purpose of this assignment Protégé, a semantic knowledge technology, was used to create a knowledge model which has completeness, connectedness, congruency, perspective and purpose.

Transformation of knowledge, contained in the construction project, to ontology started with the two technical reports (project design description document) produced by project consultants (civil engineers). The technical reports were used to identify most relevant concepts in the construction project (Fig. 6). First the two documents were compared according to their table of contents in order to align the conceptually similar chapters (i.e. description of the construction object) and their related chapters were aligned first. In the next step most relevant (and frequent) terms were identified as candidates for common concepts in the planned ontology.

These concepts were defined in a class hierarchy and the various classes and concepts were linked via object properties and data points were set for various members using the data property functionality (Fig. 7).

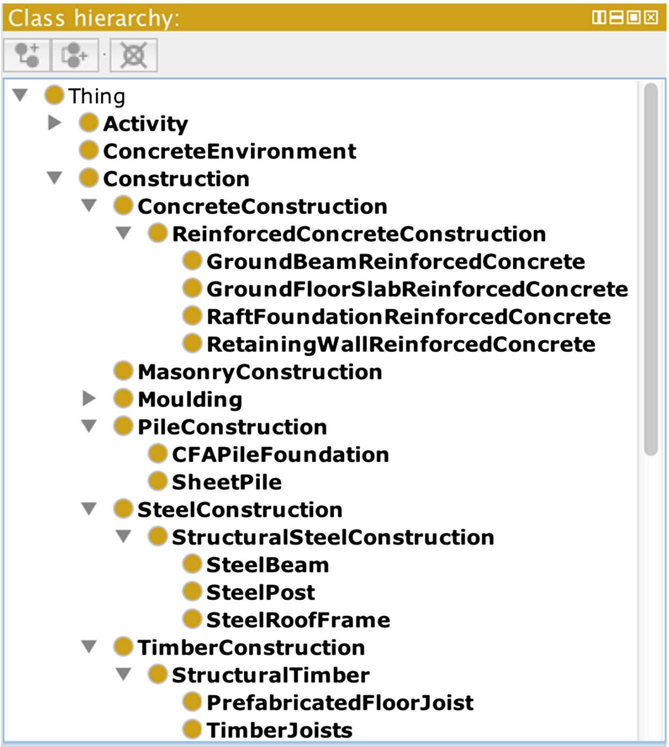


Fig. 6. Knowledge model: concepts hierarchy

Knowledge taxonomy is also a vital part of the knowledge model. A knowledge model fails if a common understanding and structure is not initially reached and understood. For the purpose of this assignment the buildingSMART Data Dictionary was used for the taxonomy, aiding in the achievement of externalisation.

The resulting knowledge model (semantic network), a construction ontology is presented on Fig. 8.

Visualisation of construction ontology shows semantic network links between various concepts created as part of the ontology. The main concepts subgroups are Project, Contract, Materials, Construction Site, Construction, Project, Design, Structural Design, type of Design and Design Codes. Figure 9 shows links created between concepts of Contract, Subcontract, Contracting Parties, Contractor and the various SubContractors. Figure 10 demonstrates the links created between the concept construction Materials and the properties of those materials.

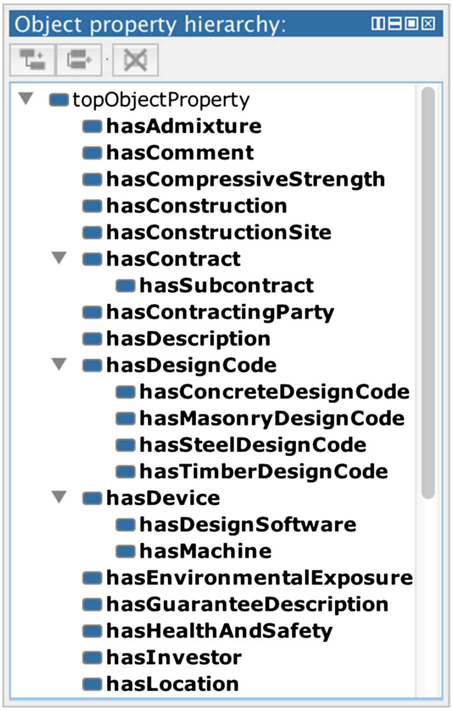


Fig. 7. Knowledge model: object properties

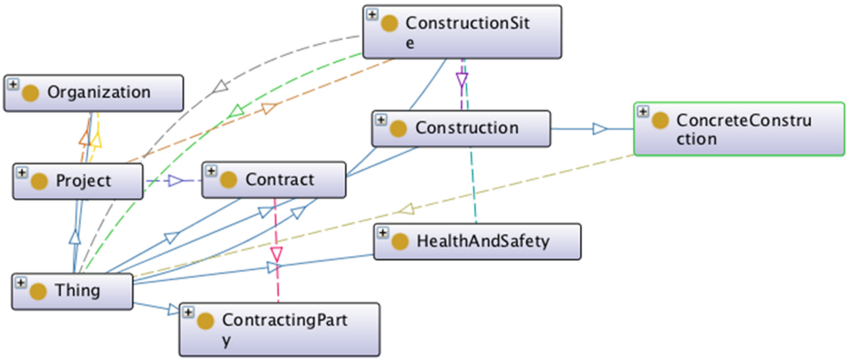


Fig. 8. Ontology graph: main concepts in the construction ontology

Analysis of the construction ontology in this development stage shows following statistics:

- 105 concepts,
- 39 object properties and
- 5 data properties.

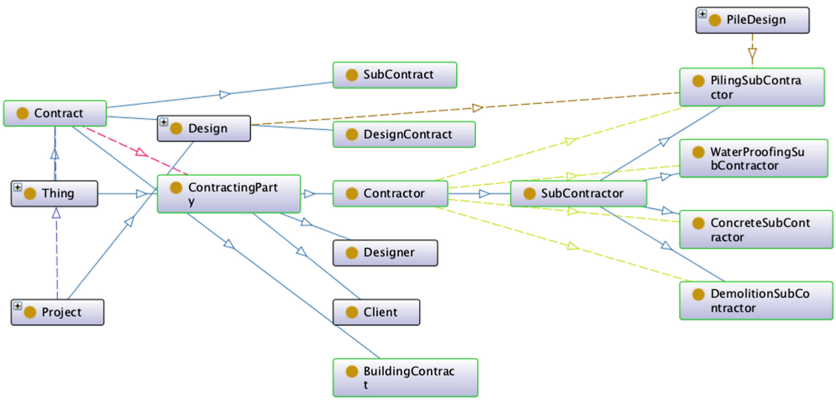


Fig. 9. Ontology graph: concept contract and its relations in the construction ontology

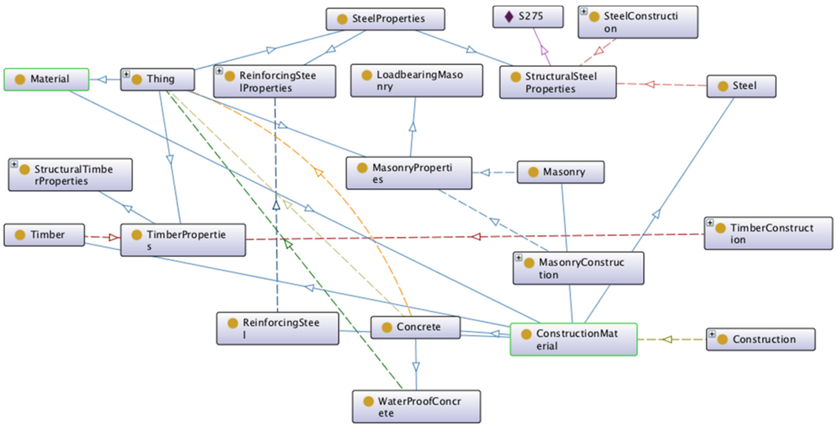


Fig. 10. Ontology graph: concept material and its relations in the construction ontology

4 Application of the Construction Ontology into Semantic Web Portal

For the knowledge management that accompanies construction process on the construction site "Underpass Grlava" we used an open source web portal MediaWiki with Semantic MediaWiki extension pack that offers a various add-ons that support semantic technologies. For instance, semantic maps extension provides a service that is capable of displaying construction sites on the map based on coordinates that were semantically annotated somewhere in the text.

Next step was integration of the knowledge model, construction ontology, which was created in Protégé, into Semantic MediaWiki. The key of this integration is to transform OWL individuals to article pages, OWL classes to category pages and OWL object or data properties to property pages. Semantic properties link articles in a

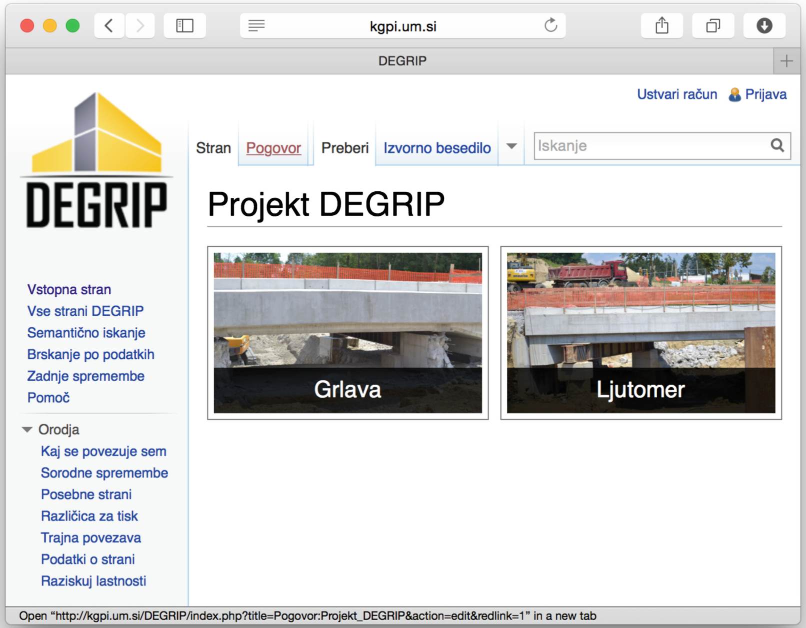


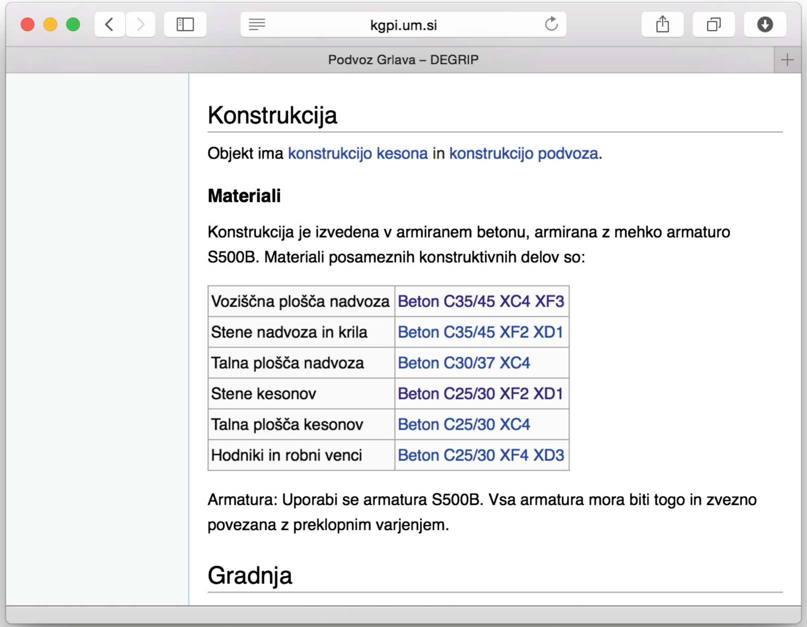
Fig. 11. Semantic web portal for construction project [10]

semantic network and, furthermore, take the plain content to the next level. User (in this case a civil engineer) gets information not only from a specific article but also from the articles that are somehow semantically connected with each other. The Semantic MediaWiki serves as a semantic portal for construction companies (from civil to design and build companies) related to the erection of infrastructure objects on the construction site "Underpass Grlava" (Fig. 11). The construction manager, site foreman and assistant site foreman use the semantic portal as a source of knowledge for daily on-site construction activities.

On the construction site "Underpass Grlava" six different types of concrete were planned for installation because the construction structure is exposed to permanent moisture due to high groundwater. One such type of concrete is i.e. concrete C25/30 XF2 XD1 where XF2 and XD1 are exposition classes used to describe chemical and physical environmental conditions to which the concrete reinforcement may be subjected. XDn denotes exposure of concrete reinforcement to chlorides (excluding seawater), XFn denotes exposure of concrete reinforcement to frost and thawing salt.

It has been realized that the characteristics of the planned concrete are not written in technical documentation available on-site. Construction site management staff admits that it is difficult to remember the abundance of information related to daily site activities.

In the construction ontology building materials are semantically related to structural elements (Fig. 12).


kgpi.um.si

Podvoz Grlava – DEGRIP

Konstrukcija

Objekt ima konstrukcijo kesona in konstrukcijo podvoza.

Materiali

Konstrukcija je izvedena v armiranem betonu, armirana z mehko armaturo S500B. Materiali posameznih konstruktivnih delov so:

Voziščna plošča nadvoza	Beton C35/45 XC4 XF3
Stene nadvoza in krila	Beton C35/45 XF2 XD1
Talna plošča nadvoza	Beton C30/37 XC4
Stene kesonov	Beton C25/30 XF2 XD1
Talna plošča kesonov	Beton C25/30 XC4
Hodniki in robni venci	Beton C25/30 XF4 XD3

Armatura: Uporabi se armatura S500B. Vsa armatura mora biti togo in zvezno povezana z preklopnim varjenjem.

Gradnja

Fig. 12. Semantic web portal for construction project: material is semantically linked to the construction

With the use of semantic technologies, which are available on construction site in the form of a semantic web portal (or knowledge portal) we can improve the quality of the construction project. The quality is improved directly (i.e. smart and semantically integrated access to explicitly written construction project information) and indirectly (better quality tacit-to-tacit knowledge exchange between site staff).

5 Conclusion

In the paper we have analysed requirements for knowledge management in the AEC industry (Architecture, Engineering, Architecture) from the notion of interrupted information flow in the lifecycle of a construction object (i.e. building or infrastructure object). The purpose of the development of knowledge model for construction projects is its integration into semantic applications. Such application is a semantic web portal, which enables simple, quick and smart access to construction project information. This information can be automatically displayed depending on the semantic links while browsing or through the user's custom semantic queries. The semantic links facilitate the running of queries. Results of semantic queries are not limited to text only format (i.e. query returns list of material for construction object), but can also return semantically generated maps (i.e. locations of construction site objects on a map) that leads to further semantically information mining.

The knowledge model results in better management of knowledge and facilitates intermediation on a construction project, where the ontology is used to facilitate knowledge sharing and reuse among all personnel involved in the construction project.

This cognitive process allows for the streamlining of cross functional decision making at all levels of project delivery. The semantic web portal is a vast repository of construction project information where the project can be described to a level of granularity as desired by the design team, which greatly facilitates knowledge transfer and sharing within a construction team.

Our future research will also seek for supporting ideas in the Internet of Things (IoT) domain. This is because IoT also depends on semantic modelling, which provides a potential basis for interoperability among different systems and applications [14]. In the construction industry IoT directs research towards automated processes during design, engineering, erection and maintenance of a construction.

References

1. Anumba, C.J., Wang, X.: Mobile and Pervasive Computing in Construction: An Introduction (2012)
2. Beetz, J., et al.: bimserver. org–an open source IFC model server. In: Proceedings of the CIB W78 2010: 27th International Conference–Cairo, Egypt, pp. 16–18 (2010)
3. buildingSMART: Standards buildingSMART. http://www.buildingsmart.de/bim-know-how/standards
4. Dave, B., Koskela, L.: Collaborative knowledge management - a construction case study. Autom. Constr. **18**, 894–902 (2009)
5. Eastman, C. et al.: BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers and Contractors (2011)
6. Flyvbjerg, B. et al.: Underestimating costs in public works projects: error or lie? J. Am. Plan. Assoc. (2007)
7. Gallaher, M.P. et al.: Cost Analysis of Inadequate Interoperability in the US. Capital Facilities Industry. NIST (2004)
8. Haron, A., et al.: Building information modelling in integrated practice. In: 2nd Construction Industry Research Achievement International Conference (CIRIAC 2009) (2009)
9. ISO 16739:2013: ISO 16739:2013 - Industry foundation classes (IFC) for data sharing in the construction and facility management industries. http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=51622
10. Jakoša, D., Tibaut, A.: Project DEGRIP home page. http://kgpi.um.si/DEGRIP/index.php/Projekt_DEGRIP
11. Jung, Y., Joo, M.: Building information modelling (BIM) framework for practical implementation. Autom. Constr. **20**, 126–133 (2011)
12. Kanapeckiene, L. et al.: Integrated knowledge management model and system for construction projects (2010)
13. Tibaut, A. et al.: Design of a Family House (2014)
14. Wang, W., et al.: Knowledge representation in the internet of things: semantic modelling and its applications. Autom. J. Control. Meas. Electron. Comput. Commun. **54**(4), 388–400 (2013)

A Novel Approach to Generating Test Cases with Genetic Programming

Sašo Karakatič(✉) and Tina Schweighofer(✉)

Faculty of Electrical Engineering and Computer Science, University of Maribor, Smetanova 17, 2000 Maribor, Slovenia
{saso.karakatic,tina.schweighofer}@um.si

Abstract. Part of the automating software testing procedure includes the automation of test cases. Automation can lower the cost and effort and at the same time can increase the quality of test cases and consequently the testing procedure. Many different approaches for test case generation are available: generation from code, formal methods and different models, among others also from UML diagrams, more precisely from UML activity diagrams. Researchers use different techniques, of which genetic programming (GP) is very popular and was used in our research. In the proposed approach we generated test cases from the UML activity diagram, from which we constructed the binary decision tree structure, which is used as an instance in the evolution process of GP. The default tree structure is used throughout the whole evolution process, only the content (the testing parameters) of the nodes changes. The process of evolution consists of several genetic operators, such as selection, crossover and mutation. The main novelty of our method is a different fitness function than we can find in existing literature. In contrast to related work where the coverage is used - we used the error occurrence for our metric. The proposed method is demonstrated on the example of an automated teller machine (ATM), where we show how the full automation of test case generation and testing is a major advantage of our method.

Keywords: Genetic programming · Software testing · Activity diagram · UML · Test cases

1 Introduction

Software testing represents an important and crucial phase in the development process. Efficient testing is necessary to produce reliable systems [1] and as surveys have revealed, more than 50 % of software development effort is spent in the testing phase [2]. The percentage, transformed into time and cost, increases together with the complexity of the developed software products. Thus, the automatization of the testing process is becoming increasingly important. Automatization can reduce costs and effort, reduce the number of human errors and faults, and thus increase the reliability of testing [3,4].

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 260–271, 2015.
DOI: 10.1007/978-3-319-21009-4_20

One of the most important parts of testing automation is automatic test case generation. As the definitions say, test cases usually consist of conditions, variables and a set of sequence actions to determine if a tested system is working according to defined specifications [2,5–7]. Test case design take a significant amount of time, so automation can be of great assistance [2]. Developing and maintaining test cases manually is relatively expensive, also because of the deep domain knowledge that is required [7]. Automated test case generation can support the increased flexibility of system product development, an immediate response to changing requirements and also constitutes a foundation for the automated code generation of functional behavior and test case execution [8].

Test cases can be generated from different sources: from code, graphs, formal specifications and different models [4]. Generating test cases from models is also known as model based testing (MBT). One of the most frequently used models for test case generation are UML diagrams [9]. Although UML is primarily a language for modeling object oriented software [10] it is also an important source of information for test case design [11]. UML diagrams are divided into two bigger groups, structural and behavioral diagrams, which partly includes UML activity diagrams, which are frequently used for test case generation. It describes the workflow and extracts the core idea from the flow charts [12]. Its main advantage is its simplicity and ease of understanding with regard to the flow logic of the system [13].

There are many different approaches of test case generation from UML diagrams. Some of these approaches are also based on genetic algorithms (GA), that have been, in recent years, a frequently used technique in software testing procedure [14]. A specialization of GA is genetic programming (GP) which is used in our test case generation method. In this paper, we propose a novel approach for automated test case generation using genetic programming, where we transform UML activity diagrams to template test cases and automatically change it through the evolution process of the GP algorithm.

The paper is structured as follows. In the next section, related work in the area of test case generation from UML diagrams and GA will be presented. After this, a proposed genetic programming method is described and is supplemented with an example of an automated teller machine (ATM), which will demonstrate the advantage of our proposed method in automated systems for testing. We finish with the concluding remarks of our work and plans for the future development of our proposed method.

2 Related Work

Many different approaches are presented for test case generation from many types of UML diagrams. Usually, test cases are generated from UML state machine diagrams, activity diagrams, sequence diagrams or from the combination of more than one UML diagram, using different techniques, different intermediate representations and producing different outputs. Different techniques were researched in the work [15]. They classified approaches as: testing based on formal specifications, graph- and tree-based methods, direct UML specification processes and heuristic approaches, including GA approaches.

Test case generation, based on GA, tries to imitate natural evolution. Many different approaches, using different UML diagrams, have recently been researched. In the work [16] GA technique for generating test data from a UML state machine diagram is presented. Input to the procedure is a sequence of triggers that can change the states of UML diagram. The sequence of triggers is used as a GA chromosome and number of fired transition is the fitness value of each data. After this crossover and mutation operation are performed, a new sets of chromosomes is created. The quality of the test data is measured by the number of transitions fired by the input and with the optimization technique - GA, they try to find the best sequence of triggers to cover the most transactions of the diagram. The approaches [17,18] also describe approaches for generating test data with GA from a UML state machine diagram. This approach is interesting, but the way it measures the quality can cause difficulties, as the larger test cases appear to be better in comparison to shorter ones. This is not always the case, as in testing, we try to find the most errors and not the longest string of actions possible.

The paper [3] also wants to find the most efficient final test case sequence with the use of GA. These generate state flow diagrams, which is a simplified UML state machine diagram, from raw data. For the diagram, all possible start to end paths are found and test sequences are formed from states, where each state has its own test cases. Two of the generated paths are selected randomly, a crossover operation is performed, and a new, more effective test case is generated. The final selection of the sequence is performed by a mutation process and a final sequence is obtained. This research did not use any quality measurements for generated test cases, but instead used a random selection method and just tried to generate as many cases as possible. The only criterion they used is that the test cases have to be feasible. The use case of any evolutionary algorithm is that we use them for a supervised and guided search towards good solutions and not just for generating random possibilities.

The research [19] uses GA to generate feasible test sequences. The UML State Machine diagram is presented as an extended finite state machine (EFSM) with the help of XMI. EFSM is then transferred to an extended control flow graph. The graph is then traversed and paths, that can be feasible or not, are generated. To find only feasible paths, the state of the objects needs to be checked and for this, GA are applied. Input from the diagram, which is the list of transitions, actions, guard conditions and parameters, and are encoded into GA chromosomes and the evaluation stage of GA, use the fitness function to determine the value of a particular test case. Fit instances of test cases are then passed to the next generation and the process is repeated until feasible paths are found. From the final population, based on coverage criteria, test cases are selected and converted to a Junit test case. The coverage criteria takes into account all possibilities and also which percentage of these were covered in a test case. Again we see that researchers preferred the test cases covering more cases than those that cover any unusual cases and potentially the most problematic ones.

The same authors as in the previous paper use evolutionary algorithms (EA) for test case generation, based on a UML Activity diagram in another study [20]. Similar to the previously mentioned research [19], the diagram is transformed into a direct graph, called an extended control flow graph. To use EA, they defined the problem domain in terms of chromosomes, that are sequences of genes that define paths in a graph. EA is used to produce paths from a formed graph and the basic approach is to generate a random population of chromosomes. After this, their fitness is evaluated and evolutionary operations are performed. In the end, the fitness of newly produced generations is again evaluated. The process continues until enough paths are found, or the maximum number of generations, has been reached. In the end, test cases are presented as a sequence and associated test data. Again they used the problematic coverage metric to evaluate the test cases. Next we will look into our proposed method and the way we propose to evaluate the test cases.

3 Proposed Genetic Programming Method

Genetic programming (GP) is an evolutionary inspired optimization method and is especially useful for NP problems without domain expert solutions [21]. Genetic programming uses a guided search procedure to find good enough solutions for a given problem, where optimal solutions are not possible within a limited computing time. The solutions in our case were the test scenarios, where we tested the method parameters in particular. In contrast to genetic algorithms, genetic programming uses a complex structure to represent its solutions, which in our case was the tree derived from the UML activity diagrams. In contrast to traditional genetic programming approaches, where trees or programs can be various shapes and sizes, we used a default tree structure for each tested system and used it throughout the whole evolution process.

There are a number of such solutions (trees) in the population (set of solutions), where the set has a limited population capacity. Every solution is evaluated by the fitness functions. This function prescribes a value for the quality of a particular solution. In our case we used the maximization function, where higher fitness meant better solutions, which is described later on.

The basic genetic algorithm procedure is shown in Fig. 1 and goes as follows. We start by generating the random starting population of solutions that form the starting population in our method. Despite starting with random solutions, the algorithm eventually reaches good and excellent solutions. We call this the guided evolution, which is done with the help of the fitness function (described later on). The random starting population only serves as a method to spread the search space for the solutions. After generating the starting population, the selection operator is used to pick two solutions that participate in the matching procedure. The selection process must contain some randomness, but still prefer better solutions. This process mimics the process of natural selection, where more fit individuals have more chances of making offspring and thus have more of a chance to spread their genes to the next generation. With the selection

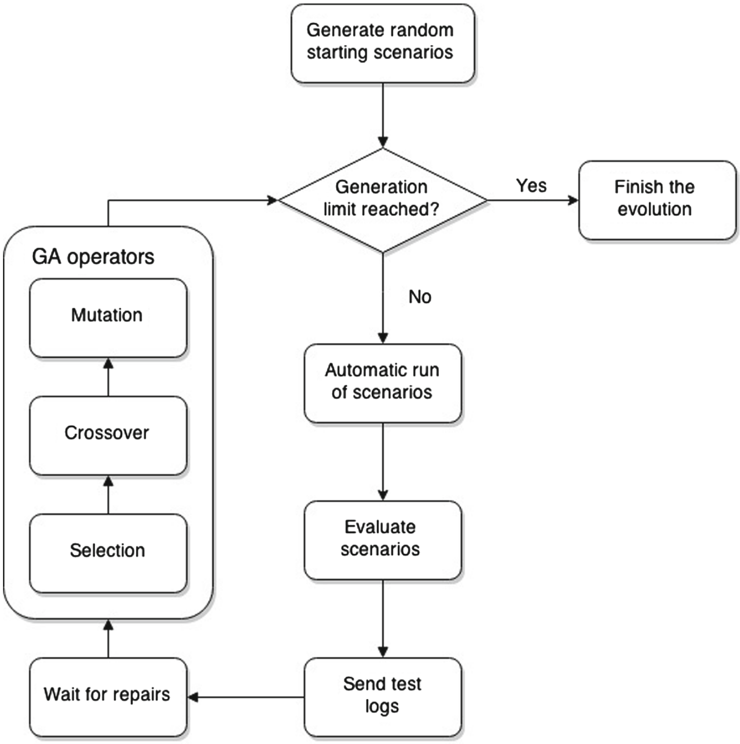


Fig. 1. Proposed genetic programming method

method we picked potential mates and then sent them to the mating process, which is called crossover in GP terminology. The crossover operator constructs a new solution from a two mating parent solution, but keeps a part of every solution in the child. This is usually done in such a way that the genotypes of both parents are mixed. After the crossover, every newborn has a chance to go through the process of mutation, again mimicking the natural process, while a mutation makes a small random change to the newborn child. This is used as the operator to expand the search space for optimal solutions and without it, solutions could prematurely converge to a local optima or not search the solution space with global optima and thus prevent any further improvements. This process of selection, crossover and mutation is repeated as long as we fill the new population with newborns, which then forms a new generation in the evolution process. With this whole process in mind, we can sometimes lose the best individual solution, because it is either not picked by the selection or its good traits are lost in the crossover and mutation processes. This is solved by elitism, whereby a limited number of the best individuals advance to a new generation automatically. This whole process is repeated until the maximum number of generations is achieved or the solutions start to stagnate and show no potential to make any improvements.

3.1 Representation of Test Scenario

Genotype is a way to encode the solutions in a way that can be used in genetic operators and, as mentioned earlier, we used the tree genotype. The novelty in our technique is to make a template, which is then used by default for all solutions. This template tree is constructed from the UML activity diagram and its specifications (see Fig. 2). The transformation is made in such way, that every actionable activity (activity that represents a method) is transformed to the tree node, and each decision unit from the activity diagram is represented as a split in the tree. The first node (the root) of the tree is the first actionable activity after the starting state of the execution and each nodes children are the next steps in the execution. Each node in the tree can have multiple children, of which the number is dependent on the number of decisions made possible by the decision units in the activity diagram.

Of course, the activity diagram can have loops, but for the sake of simplicity and with no real potential advantage, loops do not appear in our tree genotype (this would then be a diagram, not a tree). The splits of the activity diagram are also represented as multiple children, but the joins and the continuation is again not duplicated, but presented only once.

Each node in the tree has multiple properties: the name of the activity (or programming method) and the accepted parameters of this method. These parameters play a crucial role in the genetic process, as they are being changed to find potential errors in the implementation of the tested system. Parameters come with specifications that are: the type of the parameter (such as integer, real number, string of characters), and proposed boundaries for the values of the parameter.

After each generation, the solution trees were transformed to simple automated programs to test the system, which was done with the help of the Java programming language. The basic pseudo code of the transformed test case is in Fig. 3.

Each route from the root of the tree and all the way to one of the leaves is its own path in a testing scenario, which means that from each tree we get multiple sub scenarios, but common testing phases (common nodes) are not duplicated in the programming code. The results of the testing process are presented to programmers with errors in descending order based on the number of appearances, in order to emphasize the most common (and usually the most problematic) errors.

3.2 Initialization Process

Now that we have established how our genotype looks like, let us describe the initialization process of random tree creation. The type of each parameter will be used to determine which random function should be called. The boundaries are here to mark the expected values, but testing values will not only be in this set as this would not be beneficial for testing purpose. This ensures that illogical values (monkey testing) should also be treated accordingly by the programmers.

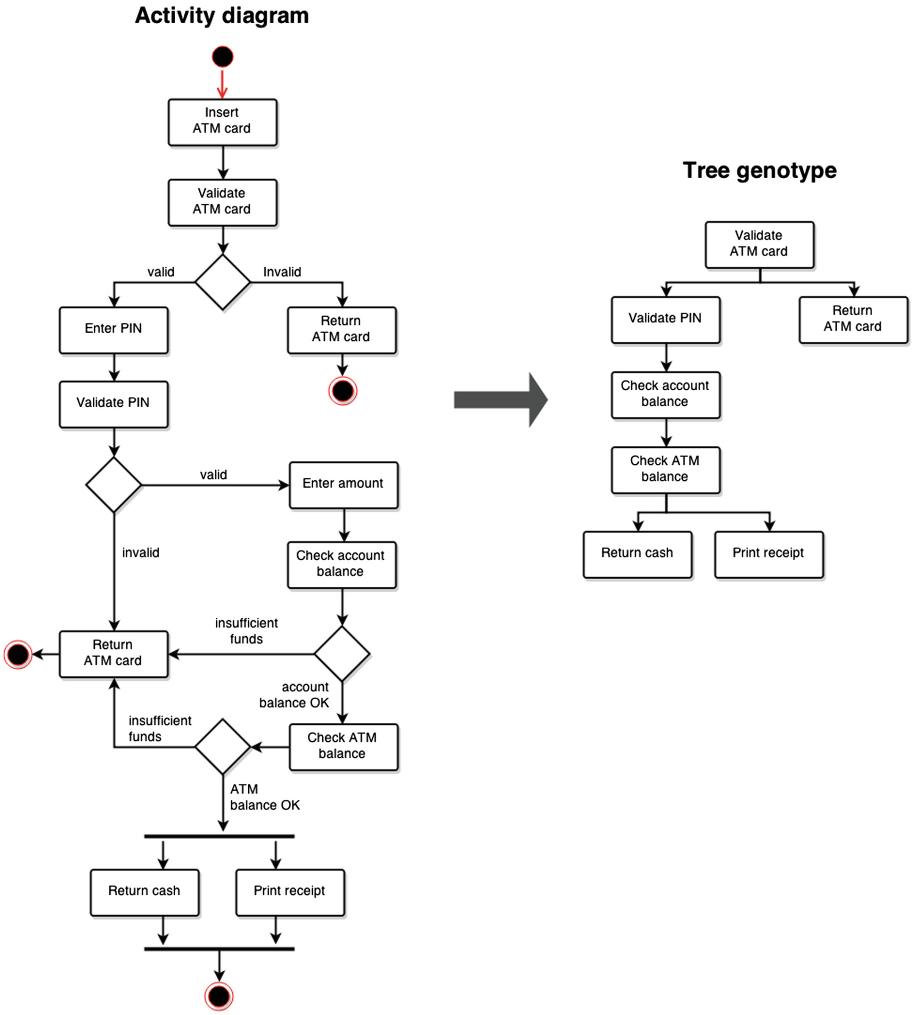


Fig. 2. Transformation from the activity diagram to the tree structure used in GP

When we have a tree structure, the random solution generation process begins. If the activity does not need any parameters or is not an actionable activity, this process is skipped and an empty node is created.

When an activity accepts parameters, they are generated randomly, but the process is different for each data type of the parameter. For integers and real numbers, the parameter is created with a normal distribution, where the mean of the distribution is the mean of the parameter boundary and the standard deviation of the distribution is the size of the boundary. This sample of normal distribution has a 68 % of chance of being in the proposed boundaries, as depicted in Fig. 4. When the parameter type is a string of characters, the

```
testScenario(Tree t, Parameters p){
  ErrorLog log

  if(t.hasChildren){
    foreach(Tree child : t.children){
      log.add(testScenario(child, p))
    }
  }

  try{
    t.executeWithParameters(p)
  } catch (Exception ex){
    log.add(ex)
    return(log)
  }
}
```

Fig. 3. Basic pseudo code of the transformed test case

algorithm chooses the string as a random length with random characters. The minimum and maximum size of parameters are considered in 68 % of cases and ignored in other cases (even empty strings appear). The same creation strategy is used in date and time parameter creation: in 68 % of cases, the parameter values fall within the expected boundaries. The boolean and enum values are chosen at random with uniform chances for each value and so are always in the prescribed boundaries.

3.3 Evaluation of Test Scenarios

The fitness function is the maximization function with the goal of finding the most errors in each test scenario: the more errors the better, but with one exemption. Fitness is not just a number of errors, where the most frequent mistakes are corrected before the not-so-common ones (as should be done), as this would result in an almost random search from generation to generation. After each scenario is executed and all errors are systematically logged, normalization ensues. Each error provided adds a value to the fitness function in accordance with the number of its appearances. So if one particular error occurs 5 times, its fitness contribution is 0.2. This way we can ensure, that common or regular mistakes are not favorably copied to the next generation even though they are most likely to be repaired in the next iteration of a program.

3.4 Genetic Operators

The crossover operator for our proposed method is a different one than those in other types of tree-based GPs. In a traditional GP, the trees do not follow

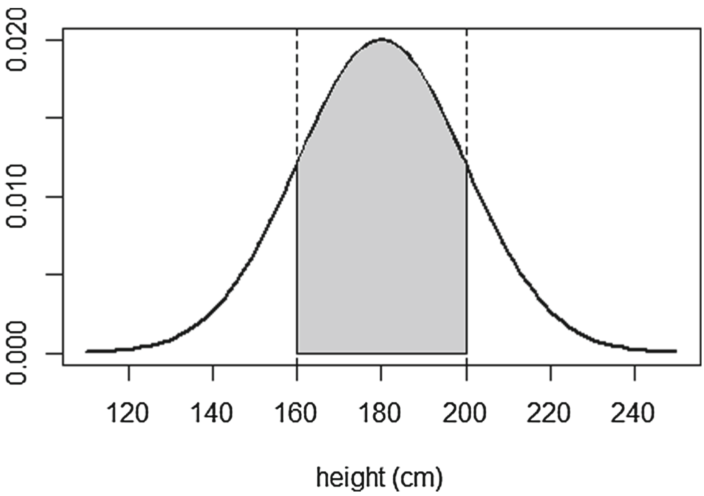


Fig. 4. Example for parameter height and the particular value probabilities

the template design but are various shapes and sizes, but in our method, the tree genotype is always the same. The consequence of this is that the crossover operator does not pick two random subtrees for the exchange, as is standard operating procedure with traditional GPs, but instead the crossover happens in the same location of the tree and it exchanges the subtrees on the same location. This has an advantage of keeping the context of subtrees and is shown in Fig. 5. The context preserving crossover has already been studied before and has proven to be better than random crossovers [22].

The mutation operator is implemented in the following way. One random node in the tree phenotype is selected and its parameter content is changed in random fashion, with consideration given to previously mentioned boundaries. Only one node is changed and not the whole subtree, as this would contribute too much of a change, which is not always welcome. For one node, all of the method parameters are changed.

4 Example

For our example, let us use one of the examples from [20],where we tested the program code of the automated teller machine (ATM) and used the simulated error handling to show the usefulness of such a GP system. The GP settings are set to the following values and were chosen as the standard default values of GP used in the literature:

- generation limit 50
- 25 solutions (test scenarios) per generation
- 100 % chance of context preserving crossover

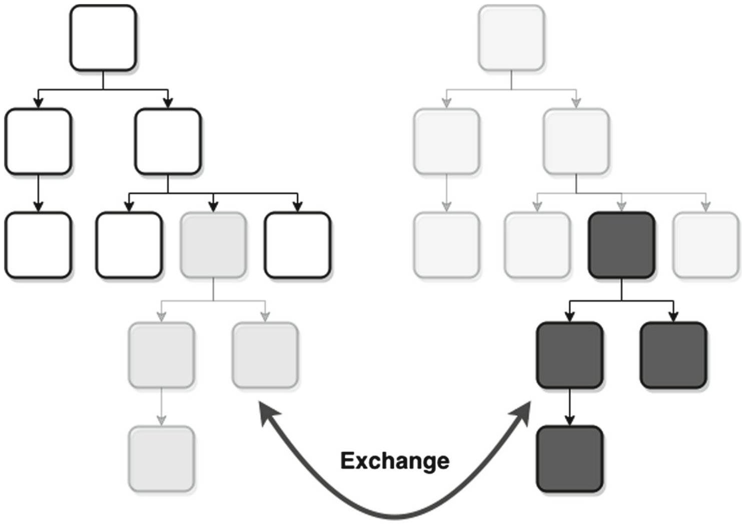


Fig. 5. Example for parameter height and the particular value probabilities

Table 1. Examples of test cases for ATM

	Validate card	Validate PIN	Check account balance	Check ATM balance	Return cash
Expected values	16 digit number	4 digit number	integer	Positive integer	Positive integer
Test case 1	7279564311512585	2119	19421	83617	8379
Test case 2	6796926169613591	0000000*	−402	65120	528620
Test case 3	7125259131937491	4965	64496	−10770∗	60387
Test case 4	5045201574473518	2727	−823	60801	−4384∗
Test case 5	3809925341311081	9738	200818	95331	6

– 10 % chance of random mutation
– 1 elite solution advances to the next generation automatically
– binary tournament selection
– maximization fitness function as described earlier ...

In the left side of Fig. 2 there is an activity diagram for our ATM and on the right side there is a corresponding tree structure that is used as an individual solution throughout the evolution. Note that in our simulation, the three most frequent errors were considered fixed in every generation. The evolution ran for 100 generations and in Fig. 4 these are plotted with a fitness value progression.

The nature of our fitness function is such that it gives more weight to less frequent errors. This is the reason why we see the gradual decrease of fitness over time when only irregular errors occur anymore.

As can be seen in Table 1, many different cases can derive from this process. The first example (Test case 1) is one test case with the following parameters for the test case: a 16 digit card number 7279564311512585, a 4 digit PIN number 2119, an account balance of 19421.00, an ATM balance of 83617.00 and 8379.00 cash returned. Note that if some parameter starts to dominate the population, it is forced out in the next generation, because it should no longer help with the testing since it should be corrected. If it happens that a correction was not made, then naturally it persists in the population with even more prevalence until its correction. Note that some values (denoted with an asterisk (*)) are out of the given expected value boundaries, which is done intentionally as explained in the previous section.

5 Conclusion

We introduced a novel approach for the automatic generation of testing scenarios, which was based on the evolutionary method of genetic programming. In comparison with other evolutionary algorithms for the construction of testing scenarios, our method uses a novel representation of a tree which proved to be a more natural fit for the main genetic operator: a context preserving crossover. Our method used the activity diagrams of tested systems to form a tree genotype to play the role of an individual in an evolutionary process of changing the method parameters. This iterative and automatic process of evolution and testing was tested on the simulated scenario of an ATM machine, where we showed the potential usefulness of our method, where it effectively stayed on the path of evolution without getting stuck in the local optima with only the most obvious errors. Our method was still not tested on a real implementation and testing environment, but this remains one of our goals and challenges for future research of the topic.

We could not compare our method to other already existing automated test case scenario methods since the standard evaluation metric of test case scenarios does not exist. This prevents the exact comparison and remains one of the challenges of this field, thus presenting a good research opportunity.

Another goal for the future is to expand the possible diagrams used with our system, for example by including the UML sequence diagrams or flow diagrams as one of the possible design techniques to form the tree genotype for test scenarios.

Acknowledgements. This work was partly supported by the Slovenian Research Agency (SRA) under The Young Researchers Programme (SICRIS/SRA code 35512, RO 0796, Programme P2-0057).

References

1. Samuel, P., Mall, R., Bothra, A.K.: Automatic test case generation using unified modeling language (UML) state diagrams. IET Softw. **2**(2), 79–93 (2008)
2. Swain, R., Panthi, V., Behera, P.K., Mohapatra, D.P.: Automatic test case generation from UML state chart diagram. Int. J. Comput. Appl. **42**(7), 26–36 (2012)
3. Gulia, P., Chillar, R.S.: A new approach to generate and optimize test cases for UML state diagram using genetic algorithm. SIGSOFT Softw. Eng. Notes **37**(3), 1–5 (2012)
4. Shamsoddin-Motlagh, E.: A review of automatic test cases generation. Int. J. Comput. Appl. **57**(13), 25–29 (2012)
5. IEEE: IEEE Standard for Software and System Test Documentation (2008)
6. Li, L., He, T., Tang, S.: Consistency checking and test generation for UML statechart diagram via extended context-free grammar (2012)
7. Schlick, R., Herzner, W., Jöbstl, E.: Fault-based generation of test cases from UML-models – approach and some experiences. In: Flammini, F., Bologna, S., Vittorini, V. (eds.) SAFECOMP 2011. LNCS, vol. 6894, pp. 270–283. Springer, Heidelberg (2011)
8. Hametner, R., Kormann, B., Vogel-Heuser, B., Winkler, D., Zoitl, A.: Test case generation approach for industrial automation systems (2011)
9. Swain, S., Mohapatra, D., Mall, R.: Test case generation based on use case and sequence diagram. IJSE **3**, 1–32 (2010)
10. OMG: Unified Modeling Language Version 2.4.1 (2011)
11. Samuel, P., Joseph, A.T.: Test Sequence Generation from UML Sequence Diagrams (2008)
12. Gantait, A.: Test Case Generation and Prioritization from UML Models (2011)
13. Kundu, D., Samanta, D.: A novel approach to generate test cases from UML activity diagrams. J. Object Technol. **8**(3), 65–83 (2009)
14. Sharma, C., Sabharwal, S., Sibal, R.: A survey on software testing techniques using genetic algorithm. IJCSI **10**(1), 381–393 (2013)
15. Shirole, M., Kumar, R.: UML behavioral model based test case generation: a survey. SIGSOFT Softw. Eng. Notes **38**(4), 1–13 (2013)
16. Doungsa-ard, C., Dahal, K., Hossain, A., Suwannasart, T.: Test Data Generation from UML State Machine Diagrams using GAs (2007)
17. Lefticaru, R., Ipate, F.: Automatic State-Based Test Generation Using Genetic Algorithms (2007)
18. Lefticaru, R., Ipate, F.: Functional Search-Based Testing from State Machines (2008)
19. Shirole, M., Suthar, A., Kumar, R.: Generation of improved test cases from UML state diagram using genetic algorithm. In: Proceedings of the 4th India Software Engineering Conference, ISEC 2011, pp. 125–134. ACM, New York (2011)
20. Shirole, M., Kommuri, M., Kumar, R.: Transition sequence exploration of UML activity diagram using evolutionary algorithm. In: Proceedings of the 5th India Software Engineering Conference, ISEC 2012, pp. 97–100. ACM, New York (2012)
21. Koza, J.R.: Survey of genetic algorithms and genetic programming. In: Wescon Conference Record, Western Periodicals Company, pp. 589–594 (1995)
22. D'haeseleer, P.: Context preserving crossover in genetic programming. In: International Conference on Evolutionary Computation, pp. 256–261 (1994)

Research About Measurability of Information Quality

Miloš Fidler and Dejan Lavbič(✉)

A.T. Kearney Svetovanje d.o.o., A.T. Kearney
(Global Management Consulting Firm), Dunajska cesta 151,
1000 Ljubljana, Slovenia
milosfidler@gmail.com, dejan.lavbic@fri.uni-lj.si

Abstract. This article will discuss ongoing research about Information Quality (IQ). Raters evaluating various IQ dimensions (accuracy, completeness, objectivity…) of same object showed low agreement level, therefore making IQ not measurable. Increase of IQ measurability to sufficient level would present an opportunity for guidelines to replace information of low with high quality. Speculations why IQ dimensions are not measurable have been made but at the same time mechanisms that improve agreement level have been proposed by researchers for validation. Moreover context in which information is being evaluated has not been yet addressed by existing research. This article will describe and explain a study that aims to create a robust model that will validate and measure effect of three different IQ aspects. Although this article is still work in progress, current results regarding research construction and preliminary testing will be presented as well as future steps.

Keywords: Information quality · Inter rater reliability · Information quality dimensions · Completeness · Accuracy · Objectivity · Consistency · Big data

1 Introduction

Information quality (IQ) remains one of the most intriguing fields of computer information science research [1, 2]. While economies, companies and our daily activities are becoming more and more data driven requirement for data and information quality is increasing. Digitalization and ongoing progress in technology established possibilities to generate and store huge quantities of data [3]. But big data by itself does not necessary also mean high quality [4]. Therefore questions: what is IQ and how are information of high quality produced became a challenge.

Arazay and Kopak in [5] state that research focusing on IQ concept, has been investigated extensively [6, 7] to identify underlying dimensions (or attributes) of IQ, such as accuracy, completeness, presentation, and objectivity which are important to information consumer. Studies [8, 9] outlined importance of perception about IQ – information consumers may perceive certain quality dimensions to be more important than are others, and for a variety of reasons, including domain expertise. Moreover Arazay and Kopak [5] proved that IQ consumers perceive IQ so differently that when being asked to rate same information (Wikipedia article), rates between consumers

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 272–281, 2015.
DOI: 10.1007/978-3-319-21009-4_21

varied so much that it was not possible to determine if information is of high or low quality. Studies [1, 10] have managed to define the basic IQ components (answering partially what are IQ components) but did not yet answer how information of high quality should be produced.

Research focus was shifted to following two questions: (1) how would IQ be made more measurable (2) and why is IQ not measurable?

If IQ is not measurable then no rule can be made to produce information of higher and lower quality. However if IQ would become measurable then information of higher quality would provide a potential to determine a rule or a standard. Understanding why IQ is not measurable could on the other hand identify key activities to improve measurability and avoid critical situations which cause measurability to be limited.

2 Measurability of IQ

Measurability challenge is common to rating systems, for example rating student's performance in University. Oakleaf [11] argued that rating system becomes objective when all raters give to same student identical or at least similar mark. Objectivity can be further increased by pre-learning coordination groups where raters rate different examples and discuss why a certain mark is most appropriate. Additionally, Okleaf calculated that agreement level varies between rating groups (librarians had overall lower rating agreement level as English teachers).

Jens-Erik Mai is in his latest research [12] confirming this argument: IQ is context-dependent, and can only be assessed and understood from within specific situation and circumstances and therefore remains centered on the meaning of information and the understanding of it in a society. Wang and Strong [7] described IQ with term fitness for use - how much consumer benefits from use of information in a specific situation. Hjorlad [12] was in his study evaluating four articles about breast cancer. He was wondering about his preferences since he was the one also rating articles. What if it would not be him assessing the articles but for instance a grade five student working on a school project? Would grades still be the same when the needs of both raters are completely different?

Eppler [13] pointed out that apart the degree to which the information at hand either meets the requirements of the particular activity in which the user is engaged (the objective view) there is also an subjective view, the degree to which the information meets the expectations of the user.

It could be that existing research measured low rater's agreement level due to "loose" definition of context in which information is rated. We are arguing that most of the research [5] was asking interviewees to rate IQ purely on hypothetical basis, giving users absolutely free hands to interpret the "why am I rating" question. This could cause a situation in which two very similar raters with otherwise high agreement level produced additionally dispersed rates due to different interpretation of context. Let's assume that two mathematical teachers with similar working experience would be asked to rate same math exams. The first would be told that his rates are purely hypothetical, he should just rate it as he feels right but second needs to rate exams right since his rates will be reviewed by national rating commission and linked to incentive scheme.

Reijers and Mendling [14] pointed out another possible reason for disagreement between raters: personal and subjective factors. In principal when adjusted to IQ topic these factors mean how a rater perceives or understands the instruction of rating. Jens-Erik Mai [12] was referencing to them as social or cultural effects, arguing that user with different interest will differently understand rating question, information and rating context. We argue that rating is a decision making process that can't be measured directly but mostly interpreted indirectly by rater characteristics.

Reijers and Mendling [14] found out that "persona" factors had higher explanatory power then model related factors. In perspective of IQ this would mean that rater's rate can be better predicted by his/hers characteristics then by the information and research context.

To summarize, to date little is known about the interrater reliability of IQ dimensions, and we can only speculate which factors effect it most. Our investigation aims to fill this gap, and our research questions which aspects effect most the rater agreement level. Moreover we hope also to calculate the overall effect of personal characteristics additional validating Reijers and Mendling [14] research at area of IQ.

3 Research Methodology

Our research focuses on underlying questions: why is IQ not measurable and how much more measurable can it become?

In order to achieve our goal we have constructed a robust model that will help as analyze three different aspects effecting measurability of IQ: rating inputs and mechanics (questioner's basic inputs and mechanics), rater's characteristics (that have potential to correlate with decision making of choosing a rate) and situation (context in which rater is evaluation information).

Each aspect will be addressed with unique approach and further clarified.

3.1 Rating Inputs and Mechanics

We are arguing that agreement level of IQ raters can be further increased as of teachers in Oakleaf [11] in a way how questions and objects in study are constructed. Therefore we have carefully analyzed rating process and defined input factors affecting ratings and mechanisms with a potential to increase agreement level.

We have defined two groups of factors and mechanisms: factors and mechanisms that were already mentioned in research but require validation (i) and purely hypothetical, not yet addressed factors and mechanisms (ii).

First group will address rater's familiarity with the topic of which objects are being rated. We will also investigate how ratings are affected with number of rated instances (learning effect) and calibration, an option to compare already given rates by objects and change them (revision effect).

Second group will address a form in which information is presented to consumer. Most of the current research [2, 5] was performed on Wikipedia articles of various size and topics. We will consider cases of rating where objects are prepared in txt only format, picture only or combination of both. Additionally, complexity of objects being rated will be varied from article summaries to simple statements. Lastly, raters rating behavior will be tracked. We will measure how much time it takes for rater to perform ratings and how many mark corrections are made.

We hope to validate rating inputs and mechanics from first group and determine the size of its effect on measurability. We expect that second group of rating inputs and mechanics will also have an effect on measurability and that effect's size can be calculated as well. Moreover we believe that both effects combined will have a significant effect on IQ measurability.

3.2 Rater's Characteristics

We will also address aspect of rater's characteristics. Rating by its nature is a decision making process where in a simplified manner different people decide differently. We hope to clarify both differences.

First, key rater characteristics with highest effect on ratings will be identified. Second, segments and groups of raters with lowest and highest agreement level will also be identified. Third, we will see which rater's characteristics have a potential to maximize agreement level within segments.

To achieve this we will try to obtain different questioner feedbacks from people of various professions (not only MBA students, teachers and librarians [5]) by leveraging our networks. Various respondents will enrich our research and will make it more applicable to real life.

Moreover we have constructed several questions that will help us identify differences between raters and open options for segmentation. Characteristics have been organized in four groups: geographic (location), demographic (age, gender, family size…), psychographic (personality, hobbies…) and behavioral (rater's relationship towards IQ).

3.3 Situation or Context

Additionally we will also upgrade existing research by setting context for raters.

Apart from existing research [2, 5] where respondents are asked about IQ purely on research basis with no further clarification of the situation in which information evaluated is used, we will also simulate two other cases.

First case will simulate how raters will behave in situation where their rating skills are being checked (to obtained golden rule). User will not only be provided with the object that is being rated but also with the golden rule standard. We hope to see how well will the rater identify and evaluate the difference from the golden rule. We believe that golden rule situations will make raters more responsible since they will have feeling that their knowledge is being checked, resulting in higher pressure and extended rating times. Agreement level should therefore increase.

Second case will simulate how raters will behave in a situation where evaluated object is used for the task at hand. We believe that tasks will make raters even more focused and their judgment more objective causing agreement level to increase. Moreover we believe that there will be a significant correlation between rater's evaluation of information quality and his success at solving task. We presume that rater's performance at tasks will significantly impact his rates.

3.4 Research Construction

A set of experiments was developed to test, validate and establish the interpreter reliability of above described aspects. Basic research settings have been unified with existing studies [5] to make results comparable as much as possible. Each research study has therefore been based on same set of assumptions (number of rating scale values, rated IQ dimension…) that have already been used in existing research.

Experiments were organized in three research studies each addressing unique situation or context (see Fig. 1).

Research I is simulating same "academic" context settings as in existing research where rater was not provided with any context apart from purely hypothetical - what he feels is the right answer. We will use this to establishing a baseline comparable to existing research (cross validation) and by also determining the significance of aspects on agreement level.

Research II is imperceptible putting rater in a situation where his knowledge is being checked. It was designed as upgraded version of research I, where effects of aspects are measured and complexity level of information to-be rated reduced. Moreover it kind of replicates real life rating situations where raters are challenged to rate randomly presented cases. We hope to validate accuracy of these systems by measuring the effect of learning on agreement level trying to found out how many examples must a rater rate to achieve certain stability. Another important validation

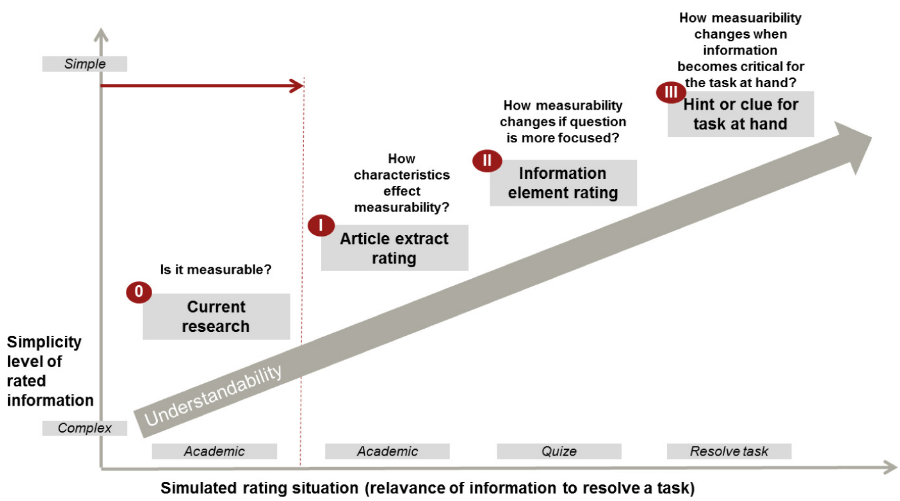


Fig. 1. Study design (self-made sketch)

mechanism will be the calibration. We will measure how selected cases and their order effect rater's requirement for additional calibration of already given rates.

Research III is further upgraded research II even more simulating real-real live examples and complexity level of information simplified. User can only rate true value of information when rated object is critical for the task at hand. We have therefore designed a special game with set of tasks. Each task consists of riddle where information is provided to a user as a hint for solution. User can then on basis of IQ and his thinking cognitive capabilities less or more successfully find a solution. Successful users are rewarded by more points, best results published as high scores and specially made storyline is motivating user not to quite. User rates the IQ of the hint before he/she tries to solve the puzzle and there is another option to change the rating after the riddle is resolved.

3.5 Research Implementation

Our study was implemented by agile approach: incremental cycles of study (product) development and testing (to collect rater requirements).

Advance research settings such as agreement mechanisms, assessed information and research mechanics were in first stage designed on common sense. We have brainstormed several possibilities and based on common sense developed study specification considering ourselves as alfa-testers.

Study was then implemented in java script as a web page questioner where beta-testing was performed by carefully selected individuals not being yet tainted with IQ research. Our crucial criteria here was the execution feasibility – is the questioner self-explanatory and short enough so that interviewee won't stop half way through. Feedbacks have been collected and implemented study further upgraded made more user friendly, easier (less clicks, more intuitive) and more fun (advance visuals) to fill.

We hope to obtain at least 300 completely solved researches of each type from Slovenian and additional 300 worldwide. Our initial target group of interviewers will be students of Faculty of Computer and Information Science. We also hope to get significant response from secondary schools. Additional interviewers will be obtained either by associations, coworkers, friends.

Apart from current research we hope to have our raters different as possible since we believe that this additionally effects agreement between them. Therefore we plan to translate the study in English and use Amazon Mechanical Turk for rater recruitment. We will also test the attractiveness level of research level III by offering it to raters as a stand-alone questioner.

4 Preliminary Results

In this chapter we will present our current survey results. First part of the chapter will focus on overall surveys performance (research I, II, III) and second on performance of rated articles in research I. Consider that this is still work in progress we will focus mostly on results of research I since we believe it is most stable. There is great change that numbers will change, since we are currently working with small sample of similar interviewees (same background).

Until this moment we have obtained 79 completed surveys for research I & II and additional 26 for research III. The difference in completed surveys occurred due to parallel roll out of research I & II, while implementation of Research III was still in progress.

Our interviewees have been students of 2nd year study of undergraduate university program performed on Faculty of Computer and Information Science, University of Ljubljana.

4.1 Survey Performance

We can say that it paid off to put extensive amount of time and effort in research design and testing due to several reasons.

First, we achieved high completion rates with minimal drops. Only 2 interviewees out of 81 have not complete Research II (98 % of completion rate). We have received 100 % completion rates for Research I & III with no drops;

Second, short survey completion times were achieved (Fig. 2). Interviewees have completed all three researches in app. half an hour (31:14 s). The largest amount of time has taken the completion of Research III with on average 19:09 s and more than 60 % of total time spent (for the interviews answering all research questions).

Third, we also managed to achieve efficient use of interviewee's time. Interviewees spent on average app. 39 s per question, making a survey very interactive (Fig. 3). Completion times where as expected progressively lengthened since each research has intentionally been designed as an upgrades of previous one to leverage interviewees learning about survey mechanics to full extent. Apart from more question, upgrades also improved context of research to be more and more aligned with real-life situations. Therefore time per question did not increase from research I to research II, since

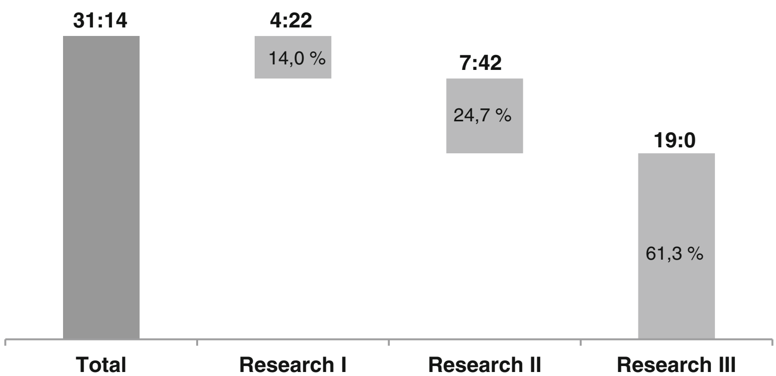


Fig. 2. Total time spent per research; Interviewees have spent on average 31:14 s to finish complete package of all three researches. The longest time to complete took research III with 19:09 s. Research II took with 4:22 s app. twice as long as research I with 7:42 s.

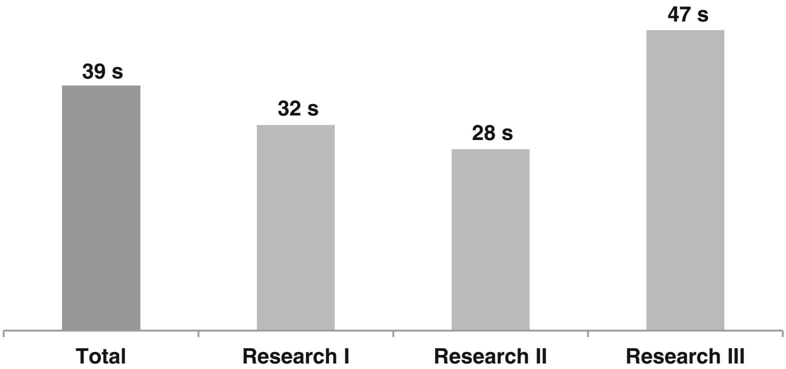


Fig. 3. Average time spent per question; Interviewees have spent on average app. same amount of time to answer a question in research I & II, app. 30 s. Time spent per question in research III exceeded time spend per question in research I & II for 60 %, resulting in 47 s. Most of the difference can be explained by interviewees solving riddles.

simulated situation did not burden interviewee with more work. However this has not been the case for research III where amount of time spend per question increased for 17 s (60 %) compared to research I & II. Interviewees have spent more time per question due to resolving riddles (simulation of task-at-hand situation).

4.2 Research Implementation

We have used in research I following articles: homo-sapiens (human), Occitan (archaic french language), alexandrite (valuable gem stone). Articles have been made as summaries of various online sources (Wikipedia, forums, encyclopedias) with same structure and length of around 180 words and 14 sentences. Each article has been structures into parts like definition, discovery, characteristic… so that topic was presented what we believed to be the most concise and efficient way. Therefore we had every reason to believe that articles will receive good marks.

Interviewees rated articles as good, they have in general agreed that overall quality of articles is slightly better than neutral (Table 1). Interviewees rated with best overall IQ mark article about alexandrite (1,28) and with the lowest overall IQ mark article about homo-sapiens (1,00). Article about alexandrite exceeded average marks in all IQ dimensions, apart from accuracy. Alexandrite was also the only article that interviewees found complete.

Overall from IQ dimension prospective interviewees rated best objectivity, giving it a very good mark. Consistency of representation received good mark, while completeness and accurate received a mark between neutral and good.

Table 1. Average grades of rater articles per IQ dimension; Raters were grading articles with 7 unit Likert, having 3 not agree, one neutral and 3 agree units. When calculated averages we mapped not agree values from -3 to -1, natural as 0 and agree values from 1 to 3.

IQ dimension	homo sapiens (human)	occitan (language)	alexandrite (gem stone)	Total
accuracy	0,85	0,28	0,70	**0,61**
completeness	0,22	0,89	1,26	**0,79**
objectivity	2,19	1,96	1,79	**1,98**
consistency of representation	0,74	1,21	1,36	**1,10**
Total	**1,00**	**1,08**	**1,28**	**1,12**

5 Conclusion and Future Works

We are implementing a three part study about IQ that will help us build a robust model and determine effects on IQ agreement level of three different aspects: rating inputs and mechanics, rater's characteristics and situation. Our study was designed in a way to add as much as possible of aspect elements neglected by current research but also to have at the same time sustainable response times. Additionally, each research faces rater with different situation in terms of usefulness of evaluated information.

Hopefully robustness of the model will help us determine how agreement level is affected by each aspect and what the scientific and practical implications are. We hope that our model will show how effects may vary in different situation for various rater segments. Moreover, we are very interested how overall agreement level (and of certain segments) will evolve throughout different researches.

At this stage we are still testing and improving our questioner. It seems that we managed to design user friendly survey with a great potential for academic contributions. We have achieved preliminary validation of rated article summaries in research I and saw that they have on average obtained slightly better mark than neutral. We hope to obtain more responses so that we are able to build IQ agreement level model and compare results with Arazay and Kopak's research [5].

References

1. Hilligoss, B., Rieh, S.Y.: Developing a unifying framework of credibility assessment: Construct, heuristics, and interaction in context. Inf. Process. Manage. **44**(4), 1467–1484 (2008)
2. Yaari, E., Baruchson-Arbib, S., Bar-Ilan, J.: Information quality assessment of community-generated content - A user study of Wikipedia. J. Inf. Sci. **37**(5), 487–498 (2011)
3. Ge, M., Helfert, M.: A review of information quality assessment. In: IET Conference Proceedings, Institution of Engineering and Technology, pp. 951–958 (2007)
4. Geiger, J.G.: Data quality management: the most critical initiative you can implement. In: SUGI 29, Montreal, Canada (2004)

5. Arazy, O., Kopak, R.: On the measurability of information quality. J. Am. Soc. Inform. Sci. Technol. **62**(1), 89–99 (2011)
6. Rieh, S.Y., Danielson, D.R.: Credibility: a multidisciplinary framework. Annu. Rev. Inf. Sci. Technol. **41**, 307–364 (2007)
7. Wang, R.Y., Strong, D.M.: Beyond accuracy: what data quality means to data consumers. J. Manage. Inf. Syst. **12**(4), 5–33 (1996)
8. Flanagin, A.J., Metzger, M.J.: The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. New Media Soc. **9**(2), 319–342 (1981)
9. Lee, Y.W., et al.: AIMQ: a methodology for information quality assessment. Inf. Manag. **40**(2), 133–146 (2002)
10. Liu, Z.M.: Perceptions of credibility of scholarly information on the web. Inf. Process. Manage. **40**(6), 1027–1038 (2004)
11. Oakleaf, M.: Using rubrics to assess information literacy: an examination of methodology and interrater reliability. J. Am. Soc. Inform. Sci. Technol. **60**(5), 969–983 (2009)
12. Mai, J.E.: The quality and qualities of information. J. Am. Soc. Inform. Sci. Technol. **64**(4), 675–688 (2013)
13. Eppler, M.J.: Managing Information Quality: Increasing the Value of Information in Knowledge-intensive Products and Processes, 2nd edn. Springer, Berlin (2006)
14. Reijers, H.A., Mendling, J.: A study into the factors that influence the understandability of business process models. IEEE Transact. Syst. Man Cybern. **41**(3), 449–462 (2011)

Product Data Analytics Service Model for Manufacturing Company

Anne-Maria Aho(✉)

Seinäjoki University of Applied Sciences, Seinäjoki, Finland
anne-maria.aho@seamk.fi

Abstract. Manufacturers are using data obtained from sensors embedded in products to create innovative after-sales service offerings. Service offers companies significant opportunities to create and capture economic value. Firms are increasingly focusing on how they can deliver services that help their customers deliver value. Service providers are shifting from being 'doers' to becoming 'problem solvers', capable of orchestrating the delivery of complex services. ABC, a product manufacturing company is moving from a product manufacturer to a product-service system (PSS). However, the shift from a product to a PPS system is not trivial. This case study shows how the use of the product data analytics service (PDAS) model can help companies who are contemplating using Big Data to provide competitive services. Understanding what is happening in the shift to services is vital to the future success of all product companies.

Keywords: Servitization · Data analytics · Big data · Product-service system (PSS)

1 Introduction

Winners in the global manufacturing world will be those companies that can adeptly harness Big Data with manufacturing analytics to uncover customer insight, identify new markets, monitor sensors and collect after sales data [1]. Exploiting Big Data through data analysis can make "substantial improvements" to how companies respond to customer needs if they can effectively gain insight from large databases and online chatter about brands or products. It is now economically feasible to embed sensors in products that can "phone home," generating data about actual product usage and performance. Manufacturers can now obtain real-time input on emerging defects and adjust the production process immediately. An increasingly important application for manufacturers is using sensor data from products once they are in use to improve service offerings. Manufacturing and production managers believe the greatest opportunities of Big Data for their function are to detect product defects and boost quality, and to improve supply planning. Manufacturing companies will also need to build the capabilities needed to manage Big Data.

To maintain competitive advantages, many manufacturing companies are moving to product service systems known as servitization of manufacturing. Servitization is now widely recognised as the process of creating value by adding services to products.

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 282–296, 2015.
DOI: 10.1007/978-3-319-21009-4_22

Service provision can offer a means of differentiation and competitive advantage for manufacturers to operate in global, highly competitive markets.

To maintain competitive advantages, many manufacturing companies are moving to product service systems known as servitization of manufacturing. Servitization is now widely recognised as the process of creating value by adding services to products. Service provision can offer a means of differentiation and competitive advantage for manufacturers to operate in global, highly competitive markets.

To maintain competitive advantages, many manufacturing companies are moving to product service systems known as servitization of manufacturing. Servitization is now widely recognised as the process of creating value by adding services to products. Service provision can offer a means of differentiation and competitive advantage for manufacturers to operate in global, highly competitive markets.

ABC is a leading specialist in machines and systems for sheet metal working. It must focus on innovation to operate at the increased pace of the global market and expectations of its customers. At the same time, it must find ways to answer demands for increased profitability to address global cost pressures while improving sustainability. In order to improve productivity and service to customers, ABC is embarking on utilizing data analytics. It also realizes the importance of adding value through customer-oriented services. Servitization is a suitable strategy for mature and stable industries, especially where physical assets have long-operating life spans such as ABC.

However, the transition from a product to a service company is challenging. There are many organizational and cultural changes needed. In addition, implementing effective data analytics for ABC also requires a new approach. To address the problems, we have developed a product data analytics service (PDAS) model to guide the transition.

This paper shows how ABC uses the PDAS model to implement its business innovation. The paper begins with a brief review of Big Data analytics, followed by PSS. The next section describes the use of the PDAS model by ABC to implement its data analytics service system.

2 Big Data

Studies show that Data-Driven Decision-Making grounded on large “volumes” of a wide “variety” of data with high-“velocity”, known as Big Data, can considerably enhance the organizational performance. Because Big Data can collect and process data sets from sensors throughout your facility, it can quickly locate flaws and defects based on data inaccuracies. This gives manufacturing companies the opportunity to solve the issue before it becomes a costly manufacturing mistake.

Big Data can track the processes of each and every sensor in your factory and allow the attached equipment to self-govern. Companies that effectively create and implement Big Data strategies stand to gain a competitive advantage.

Data from a variety of sources can be combined to improve existing products. For example, Business Week recently reported how manufacturer John Deere is combining GPS data with sensor data from onboard tractors. These tractors can be operated remotely, and can monitor crop yields while they work. Crop yield data is later used to determine precise amounts of fertilizer to deliver, according to the location in the field.

The transformative potential of Big Data lies in treating data as an asset. Extracting the most value out of data assets requires a data driven business strategy, a value-framework for data, assessing, valuating, realizing, complying, securing, and sustaining the value of data.

3 Product Service Systems (PSS)

In response to changing market conditions, manufacturing companies have traditionally become more customer-centric and innovative in a way that customers receive products that better fit their needs [2]. Manufacturing companies also have started adding more services to their total offerings as part of a differentiation strategy [3]. According to Fang et al. [4], companies with greater reliance on the service part of their business reportedly achieve better returns on sales and improve their value. Consequently, instead of services being add-ons to the product, they become the center of the total offering, with products as add-ons to the services [3]. Various terms describe this service differentiation in manufacturing firms, including service business development, servitization, service infusion, high-value solutions, and transition from products to services'.

Servitization in manufacturing has resulted in organizations offering complex packages of both product and service to generate superior customer value and thus enhance competitive edge. It is referred to as the Product-Service (P-S) transition and represents the transition between pure products to pure service offerings [5]. Within this transition exists combinations of products and services known as Product-Service Systems (PSS). Goedkoop et al. [6] defined the core of PSS as a marketable set of products and services capable of jointly fulfilling a user's need. The product/service ratio in this set can vary, either in terms of function, fulfilment or economic value.

Neely [7] defines servitization as a transformation journey - it involves firms (often manufacturing firms) developing the capabilities they need to provide services and solutions that supplement their traditional product offerings. He further suggested that servitization is "the innovation of organization's capabilities and processes to better create mutual value through a shift from selling product to selling Product-Service Systems". He said that two other definitions accompanies this: (i) the idea of a product-service system - "an integrated product and service offering that delivers value in use" and (ii) a "servitized organization which designs, builds and delivers an integrated product and service offering that delivers value in use". A well-known representative of this trend is Rolls-Royce Aerospace, which evolved from a pure manufacturer of aero engines to a (product-service) provider of aerospace solutions. Instead of selling aero engines, Rolls-Royce now contracts with many of its customers for "power-by-the-hour". In essence the customer buys the power the aero engine delivers and Rolls-Royce provides all of the support (including maintenance) to ensure that aero engines can continue to deliver power. Other examples of firms engaged in servitization include ABB, Caterpillar, GE, IBM, and Xerox [8]. Servitization is not only a prevalent trend but also that the service business is becoming an important constituent in a manufacturing firm's performance.

According to Neely [9], companies offering PSS have significantly higher revenues than companies with a product dominant logic. Bowen et al. [10] argued that offering services is a competitive advantage, which satisfies individual customer needs. The main benefits of PSS can be seen in the satisfaction of consumer demands and the continuous improvement of the business [11].

According to Gebauer et al. [3] there are two logics for understanding the transition from goods to service(s). One is based on a goods-dominant logic emphasizing value-in-exchange, in which services are viewed as a special type of good. The other is the service-dominant logic emphasizing value-in-use in the customer's business context, which considers service as a process rather than a unit of output (good) [12]. These two logics can be viewed as different but co-existing.

It is our belief that manufacturers must change their business model to provide a holistic solution to the customer, helping the customer to improve its competitiveness, rather than just engaging in a single transaction through the sale of a physical product.

4 The PDAS Model

Although PSS promises to strengthen ABC competitiveness, however, this is not easy to achieve. The main reason for this is the difficulties in changing the focus of the organization and developing the necessary capabilities for excellence in service delivery. We recommend PSS for ABC to be developed gradually on the basis of a company's core-product offering.

4.1 Case Study

ABC is a mechanical engineering company producing machines and systems for sheet metal working. It offers applications in laser processing, punching, shearing, bending, automation. ABC activities aim to give a professional, dedicated and effective support to each of their customers all over the world. It is looking into how data can help the company to improve their service to the customers. Its aim is to provide real-time, accurate information on the condition of the machine to maximize uptime and manufacturing efficiency, as well as for prolonging the productive life cycle of fabrication technology. From the data collected from ABC, the company aims to use the data to improve the service to the customers. The question is how the company can transit from a product centric to a service centric company using data analytics to improve customer service. There are many issues encountered as ABC is transforming from a product to a service oriented company. This case study shows how the PDAS model is helping the company to achieve its aims.

ABC's servitization is a transformation journey - it includes developing the capabilities need to offer services and solutions that increase their traditional product offerings. The aim is to innovate the organization's capabilities and processes to better create mutual value through a shift from selling product to selling Product-Service

Systems. It is important to educate to all the staff to change their organizational cultures and mindsets. Besides changing organizational cultures and mindsets, ABC also needs to change their operations management perspective through the use of data analytics. Meetings and training are held to convince management that adopting a service oriented approach can provide ABC strategic, customer needs and the competitive advantage. Management should be told that ABC not only needs to increase the number of services in their total offering, but should also focus on shifting the company's mindset, transforming relationships from transactional to relational and developing service offerings that genuinely meet customer needs. It is important that staffs are aware of implications of the theoretical, economical and the change of global marketing needs. Sequent sections describe the process that we have developed to achieve the change from product oriented company to a PSS company. Training is essential. It is our belief that there should be fundamental changes at ABC from top management change of mindset to organizational and cultural changes. Subsequent sections describe the PDAS model used by ABC to become a PSS company.

4.1.1 PSS Mindset

Achieving competitive advantages just through products is becoming a constant challenge for ABC. However, it is usually difficult for product-oriented companies such as ABC to exploit the potentials of services. To overcome these problems, the traditional understanding of products and services has to change and managers have to conceive the dimensions of customers' problems and providers' solutions. As manufacturers add 'service' to the body of product-centric knowledge, the tendency is to treat service activities as an extension of the body of knowledge in manufacturing and engineering.

Adopting a PSS business approach requires significant organizational changes [13, 14]. Becoming a service provider represents a shift in focus for a company since they aim to generate an increase in revenue from the service part of the business. In order to manage the shift from a manufacturing-based to a service-provider model, companies need a service strategy. Implementing a service strategy is, however, not a straight road to success. There is a risk that companies may end up in a mismatch between their organizational arrangements and their strategic market offerings [3]. We concur with Gebauer et al. [3] that the implementation of service strategies includes building up an ability to deliver services, training personnel to become service-oriented and to a certain extent developing a new organizational culture. Three workshops have been conducted for this. Participants included management of service department, service sales, management of product development and software developers. The topics covered i.e. service product system, solution sales, development trends, differences between products and services and service mindset.

4.1.2 Understanding Service Concept

ABC must understand the service-dominant logic (SDL) emphasizing value-in-use in the customer's business context [12].

Smith et al. [15] argue that one of the biggest challenges facing the P-S transition is a change in mindset from the understanding of value as that created in the production and exchange of goods, to one in which value is attained from the use of an offering aimed at achieving customer goals. The following section considers insights from the S-D logic used as a lens through which to examine this change in mindset.

Creating customer value traditionally has been focused on customer needs, satisfied predominantly through the manufacturing of products. Recently, the concept of P-S transition has increasingly evolved to value created in the function and use of the product provided rather than in its ownership. Whilst PSS recognises that customer value is achieved through use, much of its development has been based on product-based thinking. There was a lack of understanding of customer ‘needs’. This product-based thinking is often termed as a goods-dominant logic (G-D logic) [15].

Servitization in G-D logic is viewed as the phenomenon of manufacturing firms ‘adding value’ through the provision of service. It often equates the idea of ‘adding value’ to achieving higher exchange value i.e. the revenue obtained from the exchange of a product. For example, Tukker [16] suggests that by ‘adding’ value through service, the client may be willing to pay more. However, exchange value only represents one part of the value creation process in PSS [15]. Lapierre [17] shows that value created during exchange transactions represent only one level of the service value proposition, while a second level is created after the exchange is complete, that is value-in-use.

According to Vargo and Lusch [12, 18], exchange is not primarily about the exchange of goods, but the exchange of service in service dominant logic (SDL). Smith et al. [15] argue that in the exchange of service, value is achieved from the integration of skills and knowledge, known as operant resources, that operate on each other or on operand resources (such as a product) to achieve value-in-use. Consequently, whether benefits to customers are attained through tangible products or human activities, a customer-focused orientation would focus on value-in-use, delivered by the outcomes enabled by product or service activities. S-D logic proposes that a firm can only offer value propositions, and its realisation can only be through co-creation with the customer. Therefore a firm cannot ‘satisfy’ a customer; they can only collaboratively support value co-creation.

In order to help the management to understand service concepts, we conducted two workshops to introduce the principles of service and service dominant logic to them. Because the service concept was very new to ABC management, two workshops were deemed insufficient. Management needs more training in this area. Our experiences reveal that changing the mindset of management from a product oriented to a service oriented company is demanding. It is important to give adequate training to management so that they can grasp the unfamiliar concepts.

4.1.3 Adopt a PSS Business Model

To succeed in the service industry, ABC has to overcome many new obstacles. In order to overcome these hurdles in a proper way, a suitable *business model* is required. A business model describes the rationale of how an organization creates, delivers, and captures value – economic, social, or other forms of value.

Kaplan [19] describes business model with three simple “business model story elements” – value creation, value delivery and value capture. Value creation tells how the organization creates value and addresses customer experience part of the business model. Value delivery defines how the organization delivers that value, describing the operating model of the business. Finally, value capture takes into account the financial model, explaining how the organization captures the economic value for itself.

We concur with Johnson et al. [20] that there are four elements of a business model to create and deliver value. Customer value proposition refers to the value creation, whereas profit formula, key resources and key processes define the value delivery. (1) Customer value proposition includes the target customer, job to be done and offering that satisfies the job to be done. (2) Profit formula defines how the company creates value for itself. It is the blueprint of the financial aspects of the business model, including revenue model, cost structure, margin model and resource velocity. Johnson et al. stress the different aspects of the financial sustainability more than the authors mentioned earlier. (3) Key resources include resources required to deliver the customer value proposition, such as people, technology, equipment, channels and partnerships. Finally, (4) key processes together with key resources define how the value is delivered. They include processes as well as rules, metrics and norms.

In order to introduce the new service business model to the workforce, we have defined different service packages based on the features of the new service product. R&D engineers were given development and learning tasks about value delivery and value capture of the new service. They must describe the customer value produced in the different service packages. We discussed the topics in workshops. The understanding of customer value was difficult. The method for delivering the training was through teamwork, evaluation meetings, and short lectures of PSS. The participants were software developers and management of service sales.

4.1.4 Develop PSS Capabilities

To be successful in the transition from products to solutions, ABC has to develop capabilities in order to design, sell and deliver services (service capabilities) and to integrate these services into customer-specific solutions (integration capabilities). According to Bowen et al. [10] Service capabilities include establishing a service culture which, in turn, lays the foundation for increasing the degree of service orientation in the management of human resources and, more specifically, in the recruitment, development and assessment/compensation of personnel. In addition, technical expertise is required to deliver repairing, inspection and maintenance services as well as to provide design and construct services. These capabilities must also needs to be supplemented with a customer-orientated attitude, which involves listening and communication skills, essential when, for example, adapting maintenance services to the specific operational needs of customers [21].

It is our belief that the service organization should be managed as a separate organizational unit. The purposes are to gain momentum in service revenues and profits, generate a service culture and ensure a more service orientated management of human resources. This unit should takes responsibility for developing, selling and delivering services. This involves developing capabilities that can innovate services systematically, define adequate service prices and achieve superior quality. The new service unit must have capabilities that have an in-depth understanding of the customer's operational and business needs.

Other important capabilities are pricing needs, not only an estimation of the cost of delivering the service, but also capabilities to assume the operating risk of the capital goods. Adequate pricing mechanisms rely on the ability of gathering information and monitoring the usage of the product. This means IT-related skills that may be an integral part of remote services, which collect data on the status, diagnostics and usage of the products from customers' sites. Such data supports companies in assessing failure risks and predicting failure rates for capital goods.

ABC also needs to develop capabilities for understanding customer needs from a comprehensive perspective when integrating these services into customer specific solutions. Integrating a diverse set of product and service components requires multi-skilled and cross-functional competencies, which include key account management, financial expertise, technical design expertise, communication expertise and project management [22]. Creation of competence management platform includes training modules and platforms for knowledge sharing.

To help staff at ABC to develop the PSS capabilities, we conducted two workshops. The topics of workshops covered service capabilities, learning methods, the design and implementation of technical environment and sharing the responsibilities of training. Participants were from IT-management, management of service and training departments, quality management and management of software development.

4.1.5 Develop the PSS Process

Based on the core-product offering, we recommended that ABC began to develop after-sales service that can support the recovery of equipment in case of breakdown. The company should work with after-sales service to provide continuous support and guarantee uptime. In the most advanced form, ABC should take over customer activities and offer capabilities for customers to achieve outcomes – that is, provide access to resources in the form of service provision. It is best to consider services that mainly focus on existing equipment. IT is important for ABC to develop the correct processes.

To do this, we conducted interviews about the processes and documented these processes as well as modelling them. Participants for this came from different backgrounds including quality management, chief of the service product, software engineers, and management of service department.

4.1.6 Change of Customer Relationship

Identifying customer needs is an important first step in designing service offerings and operations capabilities for service delivery. PSS strategy requires ABC to change their ways of doing business. The price focused, short-term mindset entailing one-off transactions must make way for new relational strategies with single supply sources based on trust that create constant streams of revenue. Learning from the pilot projects, we found that some customers have implemented new systems. ABC was then able to collect experiences and learn from these customers. ABC was also able to get more information about customer's processes and values.

4.1.7 Developing a Services Culture

In order to deliver advanced services there is a need for manufacturers to shift to a more service orientated organizational culture. Developing a service culture is often quite difficult for manufacturers who are used to designing and building complex high-value products, e.g., talented engineers might view their future careers in product engineering rather than services. Thus, senior managers need to carefully re-position the new company focus in the minds of all stakeholders, e.g. employees, customers, shareholders and identify potential blockages in terms of processes and rewards structures that might inhibit the new culture from flourishing. Thus, designing and implementing suitable service processes becomes not just a technical issue, but one requiring the recruitment of qualified engineers with the willingness and aptitude to work in a more service-oriented, customer-focused environment. One of the important aims of ABC is to help the work force to develop a service culture by constantly reminding them of the new service concept through training and knowledge sharing during their interaction with one another.

4.1.8 Develop Data Analytics

The following sections describe how data analytics should be developed. With regard to ABC, preliminary discussion about data analytics has been conducted and three benchmarking interviews have been made. However no concrete development activities have been made yet. The planned steps in the development of data analytics are presented below.

(i) Define Business Requirements

When implementing advanced analytics with Big Data, it is critical to develop well-defined, organization-specific business requirements before starting. The most effective business requirements will have well-defined out-come oriented performance measures. It is important to ask these questions:

- What is your company trying to achieve the next 1–3 years?
- What core processes of the value chain is the strategy aiming to change?
- What key change programs are taking place?

The business strategy should be broken down into manageable pieces, (e.g. by using a simple strategy tree or Porter's value chain). It is important to start focusing the analytics strategy on the areas that really matter for business success.

As the work is still in progress, for this phase of the work, we will arrange strategy workshops to define business requirements. Some benchmarking data has been collected and preliminary discussions have been conducted.

(ii) Develop Your Analytics Vision and Set Target Analytics Maturity Levels for Your Core Processes

According to Moe [23], a business strategy is about changing one or more of the core processes of the company. In setting up the analytics strategy, we need to create a vision for how analytics should be adding value to these processes. Moe [23] argues that one way of further concretizing this vision is by using maturity models. The use of a maturity model allows the organization to have its methods and processes assessed according to management best practice, against a clear set of external benchmarks.

The maturity model should focus on the core processes of the company at a fairly high level, and using these requires two separate discussions:

1. What is our current maturity level? I.e. to what extent is the organization utilizing analytics in this very process – and are we using it in a consistent manner?
2. What should be our target maturity level? What should be our ambition for utilizing analytics and data in this process? Should we aim for automated, real-time analytics where we embed advanced analytical models into business decisions and customer-facing processes? Or should we simply aim at a lower maturity where it is up to each decision maker to utilize own-grown analytical models?

The first question is about facilitating discussions to drive a common understanding and consensus among managers and analysts. The second question should be based on facilitated discussions with the same group of people, taking into consideration e.g. strategy guidelines, market best practices, your current maturity level - and what your industry peers and competitors are doing.

(iii) Build a Big Data Capability
The most critical of capabilities is the roles, and in particular, the expertise and experience needed to devise and implement Big Data strategies. Multiple roles are needed: statisticians who are skilled in the latest statistical techniques; analysts and decision scientists who understand business measurement and experimentation and who can be the broker between statisticians and business managers; the IT group who provide guidance on selecting Big Data technologies/techniques and who integrate business intelligence tools with transactional systems such as CRM and web analytical tools. It also needs business managers and knowledge workers who own the business process and are comfortable with the new "language" of Big Data and social analytics.

ABC needs to upgrade their analytical skills and literacy. Managers at ABC must view analytics as part of the fabric of daily operations, central to solving problems and identifying opportunities. A multifaceted approach that includes training, role modeling by leaders, and incentives and metrics to reinforce behaviour is needed to adjust cultures and mind-sets. A "field and forum" approach, in which staff can participate in real-world, analytics-based training is useful so that they can learn by doing. Under almost any strategic scenario, ABC needs more analytics experts who can thrive amid rapid change.

(iv) Build a Culture That Infuses Analytics Everywhere
It is important to have employees who are passionate and skilled in exploring data and content. They must understand the implications of critical data points and apply insight to every task. ABC must engineer analytics-driven business processes and practices.

To build a transformational Big Data and analytics platform, business and IT leaders must work together to develop an effective strategy creating new roles—Chief Data Officer, Chief Analytics Officer, and Chief Data Scientist—that better address business and technology needs.

(v) Selecting the Right Tools
There are several important criteria to consider such as scalability, reliability, performance, data source consumability, and ease of deployment. Managers need transparent

methods for using the new models and algorithms on a daily basis. Terabytes of data and sophisticated modelling are required to sharpen marketing, risk management, and operations. It is important to separate the statistics experts and software developers from the managers who use the data-driven insights. The goal is to give frontline managers intuitive tools and interfaces that help them with their jobs. A Big Data and analytics platform must capitalize on real-time information that is flowing through an organization. It must capture, analyse, and correlate information as it arrives from thousands of sources. This process enables organizations to assess events as they are happening and respond with automated business processes, better agility, and improved economics.

(vi) Process and Clean Data

It is important to verify that the data matches the business goals. There several questions to ask: What are the viable proxies? Are there outliers that need to be taken into account? Does the data contain bias? Are there missing values? There are a number of methods that can be used to impute, or fill in missing values, such as mean interpolation, Kalman filter, and ARMA. Make sure that we know where the data is, who is responsible for it, and where the gaps are. The quality of the data greatly affects the analysis results. Data analytics are only as good as the data itself. This is why cleaning up data to ensure that incomplete, inaccurate, and duplicate data is removed should be the first step of any Big Data project. The first step in any analytics project is to get company data cleansed and profiled so that it can be made available for use by statisticians and other data analysts.

(vii) Identify Gaps Between Current- and Future-state Capabilities

It is significant to identify the gaps between current- and future-state capabilities:

- What additional data quality requirements for collecting, cleansing, and aggregating data into usable formats?
- What data governance policies will need to be in place for classifying data; defining its relevance; and storing, analysing, and accessing it?
- What infrastructure capabilities will need to be in place to ensure scalability, low latency, and performance?
  - How will data be presented to users? Findings need to be delivered in an easy-to-understand way to a variety of business users, from senior executives to information professionals. [24]

(viii) Develop Use Case(s)

The development process of use cases can be described as follows:

- Identify the use cases required to carry out your project.
- Map out data flows to help define what technology and Big Data capabilities are required to solve the business problem.
- Decide what data to include and what to leave out. Identify only the strategic data that will lead to meaningful insight.
- Determine how data interrelates and the complexity of the business rules.
- Identify the analytical queries and algorithms required to generate the desired outputs. [24]

(ix) Develop a Business Analytics Competency Centre

In terms of capturing and developing the right skills in the era of Big Data analytics, the creation of a Business Analytics Competency Centre that sits across the business and IT departments will be critical. This structure not only clarifies the roles and responsibilities of key stakeholders for this transformation, it also drives internal visibility, provides a mechanism for education as well as bridging the IT/business gap (and the marketing and sales teams in particular – as key individuals from these departments will need to be represented) since improving decision making amongst front-office staff will be the primary focus of these people.

Top management must be involved so that can apply institutional knowledge, navigate organizational hazards, make tough trade-offs, provide authority when decision rights conflict, and signal that the leadership is committed to a new analytics culture. Leaders should take full measure of them before assigning responsibilities or creating roles. Senior teams embarking on this journey need both to acquire knowledge of data analytics so they can understand what's rapidly becoming feasible and to embrace the idea that data should be core to their business. Only when that top-level perspective is in place can durable behavioural, changes radiate through the organization. The question that should be asked is, "Where could data analytics deliver quantum leaps in performance?" This should be happening within each significant business unit and functional organization and be led by a senior executive with the influence and authority to inspire action.

To deliver successful data analytics, ABC must have a clear strategy and well-articulated initiatives and benchmarks for success. Someone must be explicitly charged to draft a plan and give sufficient discussion or time devoted to getting alignment on priorities of the PSS. Business-unit colleagues must be quick to train mid-level managers in how to use the new model. There must be readiness and willingness to experiment in the Business Analytics Competency Centre. Managers and business analysts must be able to apply the principles of scientific experimentation to their business. They must know how to construct intelligent hypotheses and understand the principles of experimental testing and design, including population selection and sampling, in order to evaluate the validity of data analyses.

Managers should also be adept at mathematical reasoning. They should have competence in the interpretation and use of numeric data. Managers should be trained to see the Big Data picture known as data literacy. This will include competence in finding, manipulating, managing, and interpreting data, including not just numbers but also text and images. Data literacy skills must become an integral aspect of every business function and activity. Employees must understand how to protect sensitive corporate data. And leaders will need to learn to "trust, but verify" the analyses of their workforce. This knowledge must be transfer to other members of the work force at ABC.

(x) Analyse the Captured Data

The data produced is explored and visualized by patterns, trends, and clusters. Furthermore, relationships are explored and hypotheses are built according to findings. The data is then mined. Various methods can be used to facilitate pattern recognition and to organize maps for visualization such as principal component analysis, factor analysis, and multi-dimensional scaling, visualizing and communicating data is incredibly important, especially for ABC who are making data-driven decisions for the first time.

(xi) Building the Model

The goal is to produce results that lead to valuable business decisions. It is important to carefully validate the results against the initial business objective. In the next phase a Center of Excellence (CoE) can be established in order to share solution knowledge, plan artifacts and ensure oversight for projects that can help minimize mistakes. This is followed by association of Big Data with enterprise data of ABC and embedding into operational workflow/routines.

During practical implementation we will align with the cloud operating model. Analytical sandboxes should be created on-demand and resource management needs to have a control of the entire data flow, from pre-processing, integration, in-database summarization, post-processing, and analytical modelling. A well planned private and public cloud provisioning and security strategy plays an integral role in supporting these changing requirements.

(xii) Develop a Data Value Strategy

Finally, it is important to develop a Data Value Strategy which will address the entire data life cycle [25]. The Analytics Value Chain describes the process and work necessary for tactical and strategic success with digital analytics. It starts with understanding business requirements and questions, to defining and collecting data, to verifying, reporting, and communicating analytics to the next steps of optimizing, predicting, and automating from digital data using data sciences. The goal of the value chain is the creation of economic value from data analytics.

The PDAS model deals with two important new issues that ABC has to address: integration of service and the use of data analytics to improve service. To create an internal business culture that thrives on advanced data analytics technology and fact-based decision making is to start at the top of an organization. Although new analytics software and high-level executive support are a good start, they are not sufficient to foster and maintain an analytics business culture. Companies also need to make sure that their employees have the ability to make the right decisions based on information gleaned from analytics technology.

It is our belief that dedication and training are two of the keys to creating a long-lasting data analytics business culture. An analytics group with its own director could develop an analytics strategy and project plan, promote the use of analytics within the company, train data analysts on analytics tools and concepts, and work with the IT, BI and data warehousing teams on deployment projects. ABC is currently working on the transition from a product to a PSS company using the proposed model.

5 Conclusion

ABC should follow a progressive transformation from product to service. Initially, service offerings should be integrated in the pricing of products or just sold on an ad hoc basis. Next, a contract with a fixed price can cover some of the activities related to the maintenance for a specified period. At a more advanced stage and after the necessary investments, fixed price offerings may cover all service costs. Finally, the most competitive offerings are the performance-based ones where the customer payments will depend on the degree to which the provider achieves performance goals.

Big Data and analytics promise to transform the way ABC does business, delivering performance improvements not seen since the redesign of core business. These tools and techniques will open new avenues of competitive advantage. However, the development of a product data analytics service is not trivial. It requires commitments from top management, managers, and workers of ABC. Besides commitment, there should be training given to appropriate personnel as well as instilling a culture of promoting data analytics in order for ABC to succeed. ABC is currently working towards this goal.

Acknowledgements. I would like to express my gratitude and appreciation for Professor Lorna Uden from Staffordshire University. She has made a significant contribution to the completion of this research paper regarding to the development of PDAS model.

References

1. Harreld, H.: Data Analysis to Rule a New Manufacturing Era – Part 1 Spotfire Blogging Team 2014. TIBCO Software Inc. (2014). http://spotfire.tibco.com/blog/?p=17593. Accessed 9 December 2014
2. Johnson, M.D., Selnes, F.: Customer portfolio management: toward a dynamic theory of exchange relationships. J. Mark. **68**, 1–17 (2004)
3. Gebauer, H., Edvardsson, B., Gustafsson, A., Witell, L.: Match or mismatch: strategy-structure configurations in the service business of manufacturing companies. J. Serv. Res. **13**(2), 198–215 (2010)
4. Fang, E., Palmatier, R.W., Steenkamp, J.B.E.M.: Effect of service transition strategies on firm value. J. Mark. **72**(4), 1–14 (2008)
5. Pawar, K.S., Beltagui, A., Riedel, J.C.K.H.: The PSO triangle: designing product, service and organization to create value. Int. J. Oper. Prod. Manag. **29**(5), 468–493 (2009)
6. Goedkoop, M., van Halen, C., Riele, T., Rommens, P.: Product Service Systems, Ecological and Economic Basics Report 1999/36. VROM, The Hague (1999)
7. Neely, A.: What is Servitization? (2013). http://andyneely.blogspot.co.uk/2013/11/what-is-servitization.html. Accessed 11 December 2014
8. Cohen, M.A., Agrawal, N., Agrawal, V.: Winning in the aftermarket. Harvard Bus. Rev. **84** (5), 29–138 (2006)
9. Neely, A.: Exploring the financial consequences of the Servitization of manufacturing. Oper. Manag. Res. **1**(2), 103–118 (2008)
10. Bowen, D., Siehl, C., Schneider, B.: A framework for analyzing customer service orientations in manufacturing. Acad. Manag. Rev. **14**(1), 75–95 (1989)
11. Aurich, J.C., Mannweiler, C., Schweitzer, E.: How to design and offer services successfully. CIRP J. Manufact. Sci. Technol. **2**(3), 136–143 (2010)
12. Vargo, S.L., Lusch, R.F.: Service-dominant logic: continuing the evolution. J. Acad. Mark. Sci. **36**(1), 1–10 (2008)
13. Gebauer, H., Fleisch, E., Friedli, T.: Overcoming the service paradox in manufacturing companies. Eur. Manag. J. **23**(1), 14–26 (2005)

14. Baines, T.S., Lightfoot, H.W., Peppard, J., Johnson, M., Tiwari, A., Shebab, E., Swink, M.: Towards an operations strategy for product-centric servitization. Int. J. Oper. Prod. Manag. **29**(5), 494–519 (2009)
15. Smith, L., Maull, R., CLNg, I.: Servitization and operations management: a service dominant-logic approach. Int. J. Oper. Prod. Manag. **34**(2), 242–269 (2014)
16. Tukker, A., van Halen, C.: Innovation Scan for Product Service Systems. PriceWaterhouseCoopers, London (2004)
17. Lapierre, J.: What does value mean in business-to-business professional services? Int. J. Serv. Ind. Manag. **8**(5), 377–397 (1997)
18. Vargo, S.L., Lusch, R.F.: Evolving to a new dominant logic in marketing. J. Mark. **68**, 1–17 (2004)
19. Kaplan, S.: The Business Model Innovation Factory: How to Stay Relevant When The World is Changing. Wiley, Hoboken (2012)
20. Johnson, M.W., Christensen, C.M., Kagermann, H.: Reinventing your business model. Harvard Bus. Rev. **86**(12), 50–60 (2008)
21. Neu, W., Brown, S.: Forming successful business-to business services in goods-dominant firms. J. Serv. Res. **8**(1), 3–17 (2005)
22. Davies, A., Brady, T., Hobday, M.: Organizing for solutions: systems seller vs. systems integrator. Ind. Mark. Manag. **36**, 183–193 (2007)
23. Moe, C.: Six Steps to a Business-driven Analytics Strategy (2013). http://blogs.sas.com/content/hiddeninsights/2013/10/13/six-steps-to-a-business-driven-analytics-strategy/. Accessed 9 November 2014
24. Gough, C.: Capitalizing on the Power of Big Data for Healthcare; 5 Steps to Get Started (2013). http://communities.intel.com/community/itpeernetwork/healthcare/blog/2013/12/11/capitalizing-on-the-power-of-big-data-for-healthcare-5-steps-to-get-started?wapkw=identify+gaps+between+current+and+future-state+capabilities. Accessed 3 November 2014
25. Phillips, J.: Building a Digital Analytics Organization: Create Value by Integrating Analytical Processes, Technology, and People into Business Operations. Pearson, Upper Saddle River (2013)

Stable Two-Sided Matching of Slot Allocation in Airport Collaborative Decision Making

Márcio Augusto da Silva Souza[1,2(✉)], Li Weigang[2], and Alair Pereira do Lago[1]

[1] IME, Department of Computer Science, University of São Paulo, São Paulo, SP 05508-090, Brazil
{mass,alair}@ime.usp.br
[2] TransLab, Department of Computer Science, University of Brasilia, C.P. 4466, Brasilia, DF 70910-900, Brazil
weigang@unb.br

Abstract. The Airport Collaborative Decision Making (A-CDM) is a new paradigm of Air Traffic Management, which takes into account the preferences of Air Traffic Control (ATC) units and those of the airlines. This inherently leads to only satisfying the preferences of a limited number of stakeholders within the airport area. Using the matching approach for two-sided markets of Game Theory, Deferred Acceptance CDM (DA-CDM) developed an expansion of the philosophy of CDM and aggregated with the Ground Delay Program (GDP). In this paper we reformulate the DA-CDM model with a mathematic approved theory. We concentrated the manipulation of the politic mechanism in compression step of slot allocation in order to allow the user to include preferences of airport managers in addition to the ATC agents and airlines. This paper approved the stable manipulation in the DA-CDM process and also evaluated this manipulation by case study.

Keywords: Collaborative decision making · Game theory · Deferred acceptance · Manipulation

1 Introduction

In recent decades, Game Theory has been used as a mathematical theory for modeling and analysis of strategies between multiple players. One of the reasons for the success of this theory can be attributed to the diversity of actual and theoretical scenarios that it can be applied [4, 9, 16]. The Game Theory has been used to study the relationships between supply and demand of resources in societies [14].

An approach to Game Theory is the Matching theory, which aims to define, analyze and propose solutions to resource allocation problems in specific markets. Informally, a matching is a match between the players which does not violate the rules of the market. The Matching Theory for a Two-Sided Market, defines a market in which two distinct user groups that provide each other with network benefits.

In the air transport, in the case of management of airports, the runways of an airport may be classified as a limited resource aeronautics and airport infrastructure that has supply and demand for their use. The use of runways of an airport is targeted by slots.

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 297–311, 2015.
DOI: 10.1007/978-3-319-21009-4_23

Therefore, the process of allocating the aircrafts to the slots for landing or takeoff operations can be modeled as a “market”. For the arrival of a flight to occur at an airport, it should have a slot allocated at that airport [5, 11, 17].

In adverse situations, where the increase in air traffic may result in congestion at airports or air sectors scenario, the management of available resources is done by assumptions defined by the philosophy Collaborative Decision Making (CDM). These assumptions are related to concepts of property, prioritization, justice and efficiency in resource allocation. The CDM is a paradigm that tries to improve the exchange of information between the various parties involved at the airport such as Air Traffic Control (ATC), airlines and airport managers. The classical model of slots relocation and still in use by ATC is based on the paradigm of CDM, where airlines must provide reliable information in a timely manner to ATC, for a better outcome.

There are three basic steps in CDM process: Ration-By-Schedule (RBS), Substitutions/Cancellations and Compression. Last one step is re-allocation the slots by integrating the airlines with *Compression* algorithm. It fills in the blank slots, according to pre-agreed rules between the agents involved [6–8].

According to [5], the CDM model created in the 1990 seeks to relate the various affected parties so that the exchange of information between them results in better decisions in the management of air traffic. This philosophy does not consider all stakeholders in the decision-making process. Today the CDM process provides for treatment only to the interests of ATC and airlines [17].

A more current approach, called Airport CDM (A-CDM) also takes into account only the preferences of ATC agents and airlines [6]. This necessarily implies to satisfy the preferences of a limited number of stakeholders may cause lack of motivation and incentives for providing truthful information [15].

The Deferred Acceptance CDM (DA-CDM) [3] developed an expansion of the philosophy of CDM and the creation of a model for the expected GDP, and this model based on game theory, especially in the Matching Theory for two-sided markets. This paper reformulated the DA-CDM with a mathematic approved theory. It concentrated the manipulation of the politic mechanism in *Compression* step of CDM for slot allocation in order to include preferences of related airport managers as new agent in addition to the ATC units and airlines. The paper approved the stable manipulation in the DA-CDM process as a theory contribution in this topic.

This paper is structured in the following manner: Sect. 2 presents a brief introduction of game theory and Matching for Two-Sided Market. Section 3 describes DA-CDM model. Section 4 shows Manipulation mechanisms in Two-Sided Market. A DA-CDM Case Study is presented in Sect. 5. Lastly, we conclude the paper in Sect. 6 and give the directions for future work.

2 Related Work

This section presents the related work of the theoretical formulation matching of game theory and Matching for Two-Sided Market necessary for the formulation of the problem A-CDM used by [3].

2.1 Matching of Game Theory

Game theory can be defined as the theory of mathematical models that studies the choice of optimal decisions under conditions of conflict [13, 14]. The basic element in a game is the set of players who participate. Each player has a set of strategies. When each player chooses his strategy, then we have a situation or profile strategies within all possible (profiles) situations.

In matching, the preferences of each player on available resources are the strategies that will be evaluated by the allocation mechanism. The Preferences are formalized through lists where each player gives their priorities on the features they want, in order of importance.

In this market, the matching mechanism should find a satisfactory outcome allocating these players with each other, taking into account the preferences of each. One matching in this context is the final result containing the allocation or exchange of resources among all players present in these markets [16].

The matching concepts are based on the key idea that agents are rational in a market by setting their preferences according to their interests and properly acting to achieve those goals. If no agent finds a way to get a better result than the result proposed by the matching process, we say that the result is stable.

2.2 Matching for Two-Sided Market

Definition 1. Formal Definition - There are two disjoint and finite sets of agents, one called S, with n agents, another called M with k agents. Each $s_i \in S, i = 1, \cdots n$, has a strict preference, and complete transitive on the elements of the other set.

This preference, therefore, can be represented by an ordered list in the set $M \cup \{s_i\}$; denotes the list preference of s_i by $P(s_i)$. So, for example, $P(s_i) = m_1, m_2, m_3, m_4, s_1, m_5$. The notation $m_1 \succ_{P(s_i)} m_2$ indicates that s_i prefers strictly m_1 to m_2. And $s_1 \succ_{P(s_i)} m_5$ indicates that s_i prefers not to be allocated than to be allocated to m_5. An agent m is an acceptable partnership, or just acceptable to the agent s if and only if $m \succ_{P(s_i)} s$. Similarly each agent m_j in M, to $j = 1, \cdots k$ have strict preference, complete and transitive in the set $S \cup \{m_j\}$, that can be represented by an ordered list of preferences, denoted $P(m_j)$. We say s is acceptable for m if and only if $s \succ_{P(m)} m$, that is, prefer s than to be allocated. Denote $P_S = \{P(s_1), \cdots P(s_n)\}$ and $P_M = \{P(m_1), \cdots P(m_k)\}$. Therefore, $P = (P_S, P_M)$ represents the preference lists all the agents of this market.

It also has set S of each agent may be assigned to at most one set of the agent M. Each m_j can admit up to $q(m_j)$. That is, $q(m_j)$ represents the number of places in m_j. Denote q by the set $\{q(m_1), q(m_2), \cdots, q(m_k)\}$. The market can be denoted simply by $H = (S, M, P, q)$ and the function of H is to find the best way of partnering between the agents. A set of such partnerships is called matching.

Formally, a problem of matching is a 5-tuple

$$H = (S, M, P_S, P_M, q) \tag{1}$$

Thus it proposes the incorporation of another element in this 5-tuple another element, the element r, which is a constraint on the allocation. Thus we have:

$$H = (S, M, P_S, P_M, q, r) \tag{2}$$

Definition 2. A matching μ is a mapping from the set $M \cup S$ into the set of all subsets of $M \cup S$ such that: (i) $|\mu(s)| = 1$ for all $s \in S$ such that if $\mu(s) \notin M$ then $\mu(s) = s$; (ii) $|\mu(m)| = q(m)$ for all m, and if the number of elements in $\mu(m)$, say r, is such that $r < q(m)$, then $\mu(m)$ contains $q(m) - r$ copies of m; (iii) $\mu(s) = m$ if and only if $s \in \mu(m)$.

Definition 3. A matching μ is individually rational if every pair (m, s) such that $\mu(s) = m$ is acceptable to s, and s is acceptable to m. In other words, the matching μ is individually rational if no agent is matched to an unacceptable mate.

Definition 4. A matching μ is stable in pairs if it is individually rational, and there is no s and m associated by this matching, $m \succ_{P(s)} \mu(s)$ and $s \succ_{P(m)} s'$ such that for some $s' \in \mu(m)$.

If there is a pair (m, s) then we say that this destabilizes the matching μ, because it can be improved if they form a partnership with each other. A matching μ is said to be unstable by sets, if there are another matching μ' and a coalition $A \subset M \subset S$.

Definition 5. A stable matching of sets if not unstable by sets. That is, if it is not destabilized by any coalition.

Stable matching was proved by Gale and Shapley [10] through an algorithm starting from the preferences and coordinates of agents that produces a stable matching in a finite number of steps.

Definition 6. The matching μ is *not blocked by pairs,* if every pair (m, s) with $s \succ_m \mu(m)$ it has to be that $m' \succ_s m$ for all and if $\left|\mu^{-1}(s)\right| < q_s$ then $m' \succ_s$.

Definition 7. The matching μ is blocked by a pair of agents (m, s) if they prefer each other to their current mates i.e. $s \succ_m \mu(m)$ and $m \succ_s \mu(s)$.

Where $\succ_p$ is the preference relation of agent p. The pair (m, s) is called a blocking pair. Now we have two criteria for a stable matching.

Definition 8. The matching μ is stable if it is individually rational and there are no blocking pairs.

Definition 9. If for every pair (s, m), $m \succ_S \mu(s)$ implies that $\left|\mu^{-1}(m)\right| = q_m$ we say that there is no waste in matching μ.

Stability is a very demanding concept from both a theoretical and empirical point of view. On the other hand, in Air Traffic Management, the safety is the maximum requirement for whole system and all time [5–8, 17]. For the slot allocation in A-CDM, the stable solution is important issue. The existed research [3, 15] developed models for modifying the Compression step.

3 The DA-CDM Model

The DA-CDM model [3] has been developed based on the Deferred Acceptance allocation mechanism from [10] and especially for CDM [15]. Our research is motivated by the maturity of this model and its application in a wide range of scenarios. The *Compression* step of CDM that we study can be seen as a problem of resource allocation, characterized by a two-sided market with on one side flights (airlines) and on the other side slots (airport managers). Implementing a DA approach in this context aims at assigning flights to slots based on a one-to-one relationship while respecting the preferences of each allocation.

3.1 Market of Slot Allocation

The problem of market of slots consists in the following elements:

i. A finite set of flights $F = (f_1, f_2, \cdots, f_n)$;
ii. A finite set of slots $S = (s_1, s_2, \cdots, s_n)$;
iii. An array of capabilities $q = (q_{s_1}, q_{s_2}, \cdots, q_{s_m})$, where q_s is a positive integer indicating the maximum number of flights that can be allocated;
iv. A list of preferences $\left(P_{f_1}, P_{f_2}, \cdots, P_{f_n}\right)$, where P_F is the relationship of flight references f in relation to slots; including the option not to allocate, $S \cup \{\emptyset\}$;
v. A list of preferences of slots in relation to flights (individually) $P_S = (P_{s_1}, P_{s_2}, \cdots, P_{s_n})$ where P_S is the slot's preferences c in relation to flights including the option to keep the position open, $F \cup \{\emptyset\}$;
vi. A restriction to the timetables of flights e_f (The Earliest Possible Arrival Time - EPAT), where $e_f \leq s_i$, i.e. any flight can be allocated a time slot that has less than the original flight schedule.

Thus we have $\mu = (S, C, P_f, P_s, q, e_f)$, which can be rewritten as:

$$\mu = (S, C, \succ F, \succ S, q, e_f) \tag{3}$$

Having defined the problem, we wish under any criterion to allocate the flights in the available slots. This allocation is referred to as a matching and its formal definition is the following.

Definition 10. A matching is a function $\mu : F \to S \cup \{\emptyset\}$ satisfying $\left|\mu^{-1}(s)\right| \leq q_s$ for all $s \in S$.

Consider a matching μ and a pair (f, s) such that $\mu(f) = s$. If the pair is not mutually acceptable (i.e., if s is not acceptable to c, or c is not acceptable to s) say that the matching μ *is blocked unsatisfied by side*. So if flights and slots are free to refuse the proposed matching, we have a first desirable property.

3.2 Agent Selection

We consider three types of agents as decision partners in the CDM process [3]:

- ATC - characterized as a single agent responsible for detecting congestion in advance by predicting aircraft occupancy in the air scenario, using data available from the flight schedule. Its goal is to control and optimize air traffic flow.
- Airline - characterized as agents having flights that will be operating during a given day. Each agent's goal is to control its aircraft with regard to planned times of take-off and landing, reporting possible schedule changes due to technical and/or mechanical problems, or cancellations that may interfere in the original flight schedule.
- Airport - represented by the airport managers of origin and destination, defined in the flight schedule. Their goal is to maintain the appropriate flow of take-off and landing in their runways, adapting to the operational security restrictions specified by the ATC agent.

It should be noted that the ATC agent represents a centralizing agent in the market and has no preferences over allocations. The Airline and Airport agents are the decision-makers within this scenario, and are responsible for determining strategies based on their own goals, thus enabling the correct formulation of a new schedule for the use of airport runways.

3.3 Reward Structure

As an initial proposal [3], a simple approach to the modeling of the objective function of the agents Airlines has been developed as the Eq. 4. In this approach, we rely on a strategy focusing on operating profit (operating profit) of each aircraft belonging to the set of a given airline flights.

$$R_F(f) = \alpha(f)\left[\left(\sum\nolimits_{k=1}^{q} sr(p_k) - vc(p_k)\right) - fc(f)\right] \tag{4}$$

where sr represents sales revenue (sale revenue), vc variable cost (variable cost) and fc fixed costs (fixed cost), per passenger p of flight f, for a total of q passengers in this same flight. The α function represents the importance given to flight f by their airline, being a value x, where $0 < x \leq 1$. In this configuration, the higher the R_F value, the better the return of the flight to the airline responsible.

The objective function of the Airport agent, given by the Eq. 4, was based on a strategy that prioritizes flights according to the amount of passengers and the time delay of the aircraft. This policy allows the decongestion of the interior of the airport of origin, the flow of traffic of people expected to the destination airport, and the reduction of stress on the crew and passengers of each flight.

$$R_S(f) = \beta(f)\, q^{\theta(t - at(f),\, c)} \tag{5}$$

where t is the current time, at the expected time of arrival (arrival), q is the total number of passengers for the flight f and c is a constant adjustment. The β function is the importance given by the airport operator to flight f, with a value x, where $0 < x \leq 1$.

Taking $D_S(f) = \theta(st(t) - at(f), c)$ representing the flight delay f and can rewrite the Eq. 5 as:

$$R_S(f) = \beta(f)\, q^{D_S(f)} \tag{6}$$

The function $\theta(f)$ processing the result of the difference between the times slot $st(t)$ and $at(f)$. If the calculation is zero or negative, indicating that the flight is not delayed, the θ function returns to a value of 1. If the calculation is positive, this value is divided by the adjustment constant c, where the value of the integer part will be returned by θ. Thus, the higher the value of c, the less importance is given to delays on flights on n. This function is based on the work of [2, 8].

It is important to note that the equations presented allow user to set a priority for the flights affected by (GDP), enabling a proper order between them.

As both reward functions are based on the number of passengers, we understand that giving priority to flights with more passengers and/or further delays will make the airport facilities, be less crowded. This reward will benefit and lead to less stress for everyone, more space, less clutter in landing and take-offs, and faster airport processes.

It is important to note that, according to the CDM process, there is no direct relationship between the priority given to a flight by the airport manager and the priority given to the trail by the airlines. The priority given by the airport manager for flights is related to the amount of passengers and the delays in order to increase passenger throughout airports facilities. The priority given by airlines for their flights is related to the costs and profits earned by each flight revenue.

3.4 DA-CDM Allocation Model

The DA-CDM model [3] has been designed for the *Compression* step after *Substitutions and Cancellations* step in CDM. The model consists of two algorithms, namely, a pre-processing algorithm and an allocation algorithm, see the detail in [3].

The pre-processing process aims to create the preference list of agents, using the Eqs. 4 and 6. After formulating the slot market for the allocation process, allocation algorithm can be used to achieve a stable matching, considering the preference of each of the market participants. The algorithm ensures correct processing in case of inconsistencies in the preference ordering of flights and slots, $\succ_F, \succ_S$.

The algorithm must always follow the preference ordering of all allocable elements in the model. Each flight is definitively allocated to the slot it has been associated to in the last step of the algorithm.

4 Manipulation Mechanisms in Two-Sided Market

A manipulable mechanism induces a non-cooperative game on the CDM, making the choice of the best strategy a very difficult decision, besides allowing the emergence of a group of sophisticated players, inducing imbalances.

The empirical research, as shown in [1, 15] indicated the existence of other forms of manipulation originated from games induced by mechanisms which may deliver other forms of strategy, not via preferences. Airline can handle a mechanism announcing an unreal number for their number of flights or performing their partnerships. In the first case, despite the impossibility of change in priorities, airline may declare a smaller quantity of flights than their actual capacity, in order to get a group of more desired flights. In the second case, a mechanism may be subject to manipulation if an airline and a flight combine a partnership out of the market.

Finally, if the priorities are defined through information provided by the airlines, a new way of handling that has not been explored in the literature can be caused by incorrect information by the airline. Our study in this topic is expected to contribute a theory proof to DA-CDM model and application.

4.1 Manipulation via Capacity

Definition 11. A mechanism φ *is non-manipulable* via *capabilities* if for every $(P_F, P_S, Q) \in \varepsilon$ for all $s \in S$ *and* for all announced capacity vector $Q^{'} = (q_s^{\prime}, q_{-s})$, $\varphi(P_F, P_S, Q) P_S \varphi(P_F, P_S, Q^{'})$.

Theorem 1. *Suppose there are at least two flights and three slots. Then, the optimal stable mechanism for the CDM is manipulated via capabilities.*

Demonstration: Consider the problem with three slots $S = \{s_1, s_2, s_3\}$ and two flights $F = \{f_1, f_2, f_3\}$. Their profiles with preferences, priorities and capabilities vector are defined as follows:

$$
\begin{array}{ll}
P(s_1) = f_2, f_1 & P(f_1) = \{s_1, s_2\}, \{s_1, s_3\}, \{s_1\}, \{s_2, s_3\}, \{s_2\}, \{s_3\} \\
P(s_i) = f_1, f_2 & P(f_2) = \{s_2, s_3\}, \{s_1, s_3\}, \{s_3\}, \{s_3, s_2\}, \{s_2\}, \{s_1\} \\
P(s_3) = f_1, f_2 & Q = (q_1 = 2, q_2 = 1)
\end{array}
$$

By Gale and Shapley mechanism [10], we have the following allocation:

$$
\varphi(P_F, P_S, Q)(f_1) = \{s_2, s_3\} \qquad \varphi(P_F, P_S, Q')(f_2) = \{s_1\}
$$

But if f_1 states only one position available, i.e. $Q' = (q_1 = 1, q_2 = 1)$, it has been:

$$
\varphi(P_F, P_S, Q')(f_1) = \{s_1\} \qquad \varphi(P_F, P_S, Q')(f_2) = \{s_3\}
$$

Note that:

$$
\varphi(P_F, P_S, Q')(f_1) = \{s_1\} P_{s_i} \varphi(P_F, P_S, Q)(f_1) = \{s_2, s_3\}
$$

Therefore, φ is manipulable via capabilities.

According to [12], the assumption of compatibility between the preferences of slots and priorities of airlines guarantees non-manipulability of the TTC mechanism.

4.2 Manipulation via Priorities

Another way of manipulation that has not yet been explored in the literature is connected to the priorities formulation process [12].

Let $\mathfrak{J}_s$ the class of all possible priorities profiles. Define $\Delta_I(\Delta(i_1), \cdots, \Delta(i_{n_1}))$ as the vector information associated with the set of slots. Δ is the set of all vectors of information. This information qualifies slots, being necessary to define their planning priorities of airlines.

Thus, given a particular legal rule and a finite set of slots, a profile of priorities for airlines is determined: $P_s = (P(s_1), \cdots, P(s_m))$.

Definition 12. Ω is the mapping between the vector information of the slots and $\Im_s$

$$\Omega : \Delta \to \mathfrak{J}_I \quad \Omega : (\Delta_I) \to P_s$$

With this, a new possibility of manipulation arises, depending on the information as Δ_i provided by the slot, the generating process priorities Ω can provide a profile that benefits the airline.

Definition 13. A mechanism φ is *non-manipulatable* via priorities for all $(P_I(\Delta_I, \Omega), Q) \in \varepsilon, i \in \varepsilon I$ and $\left(\Delta_i^* \neq \Delta_i, \varphi(P_I, (\Delta_I, \Omega), Q\right)(i) P_i \varphi(P_I, \left(\Delta_i^*, \Omega\right),) Q(i)$.

5 DA-CDM Case Study

In this section, the case study uses the air movements of arrival of Tancredo International Airport (SBCF) in the city of Belo Horizonte, Minas Gerais, Brazil. This is due to a large number of daily usages of various national and international flights, where its capacity is about 10.2 million passengers per year [3].

5.1 Simulation Scenarios and Data

The SBCF airport only has one runway and the original schedule of flights is usually adapted allowing arrivals and departures. This feature allows us to test the settings of the airlines on the possible arrival time (e_f), in other words, the proposed model allows airlines to strategically advance or delay their flights, while operating restrictions are allowed. Therefore, airlines can prioritize some flights over others using the results from the Eqs. 4 and 6. To resolve this issue some additional data from the National Civil Aviation Agency (ANAC), INFRAERO, TAP Portugal, AZUL and GOL. The Tables 1 and 2 shows the flight movements of the SBCF.

As the results presented in [3], after performing pre-processing algorithm and the allocation algorithm has the following process result shown in Table 3.

5.2 Manipulation of Simulation in the DA-CDM Process

As the main part of the work, we assume that with the lack of communication of the airline to airport managers, i.e., the flight f_6, was canceled. We have the

Table 1. Air movements of arrival from SBCF [3]

Flight ID	Flight number	Origin	Aircraft	Passenger capacity	Allocated slot
F1	TAP-0101	LIS	A332	268	10:28 pm
F2	AZUL-2557	SDU	E190	110	10:32 pm
F3	AZUL-2418	GRU	E190	110	10:35 pm
F4	AZUL-4190	VCP	E190	118	10:46 pm
F5	GLO-1091	BSB	B378	183	10:49 pm
F6	GLO-1670	SDU	B378	183	10:55 pm
F7	GLO-1320	CGH	B378	183	10:58 pm
F8	AZUL-4952	CWB	E190	118	11:14 pm

Table 2. Information of flights

Flight ID	Flight number	Passenger capacity	Ticket per passenger	Costs per flight
F1	TAP-0101	268	$2,100	97.6 %
F2	AZUL-2557	110	$230	97.6 %
F3	AZUL-2418	110	$170	97.6 %
F4	AZUL-4190	118	$165	97.6 %
F5	GLO-1091	183	Canceled	Canceled
F6	GLO-1670	183	$155	97.6 %
F7	GLO-1320	183	$310	97.6 %
F8	AZUL-4952	118	$250	97.6 %

Table 3. DA-CDM process result

Slot	Flight	Owner	e_f
s_1	f_1	TAP	1
s_2	f_3	AZUL	2
s_3	f_2	AZUL	2
s_4	f_7	GOL	4
s_5	f_6	GOL	5
s_6	f_8	AZUL	6
s_7	f_4	GOL	4
s_8	Empty	AZUL	

following situation resulting process DA-CDM, with our goal to find a better allocation (Table 4).

Using data recovered from the sites mentioned the pre-processing algorithm calculates earnings per flight using the payoff function described in Eq. 4.

The reward function $R_F(f)$ shown in Table 5 results in a rank of high priority for all the flights affected by delays in the program on the ground. These results highlight the most profitable flights for the airlines.

Table 4. DA-CDM process result modifie

Slot	Flight	Owner	e_f
s_1	f_1	TAP	1
s_2	f_3	AZUL	2
s_3	f_2	AZUL	
s_4	f_7	GOL	4
s_5	Empty	GOL	5
s_6	f_8	AZUL	6
s_7	f_4	GOL	
s_8	Empty	AZUL	

Table 5. Results for applying Eq. 1 on air movements' data

Flight ID	Occupancy rate	Sales revenue	$R_F(f)$	High-priority order for flights
F1	80.7 %	$454,180	$10,900.31	1
F7	80.6 %	$45,724	$1,097.39	2
F8	80.6 %	$23,777	$570,65	3
F2	80.6 %	$20,392	$489.40	4
F4	80.6 %	$15,693	$376.63	5
F3	80.6 %	$15,072	$361.73	6
F6	Canceled	Canceled	Canceled	0
F5	Canceled	Canceled	Canceled	0

Now this information can be used to create a preference ordering for each of its flights. The preference ordering of fights $\succ \boldsymbol{F}$ is available in Table 6.

Table 6. Preference ordering of flights F

Flight ID	Order for flights	Order for airline	Adjusted ef	$\succ \boldsymbol{F}$
F1	1	1	10:28 pm	$s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8$
F8	3	1	11:14 pm → 11:08 pm	s_6, s_7, s_8
F2	4	2	10:32 pm	$s_2, s_3, s_4, s_5, s_6, s_7, s_8$
F4	5	3	10:46 pm	s_4, s_5, s_6, s_7, s_8
F3	6	4	10:35 pm	$s_2, s_3, s_4, s_5, s_6, s_7, s_8$
F7	2	1	11:14 pm → 11:08 pm	s_4, s_5, s_6, s_7, s_8
F6	0	0	Canceled	Canceled
F5	0	0	Canceled	Canceled

In this example, flights f_8 and f_7 were considered most profitable for its airlines and were prioritized over others flights. The e_f was adjusted for flight f_8 from 11:14 pm to 11:08 pm, enabling it to run for an earlier slot in allocation process. The same was done with flight f_7 and its new e_f, from 10:58 pm to 10:52 pm.

Table 7. Original and new schedule for each slot

Owner	Slot ID	Original slot schedule	Slot schedule after RBS
TAP	s_1	10:28 pm	10:28 pm
AZUL	s_2	10:32 pm	10:36 pm
AZUL	s_3	10:35 pm	10:44 pm
AZUL	s_4	10:46 pm	10:52 pm
AZUL	s_5	10:49 pm	11:00 pm
GOL	s_6	10:55 pm	11:08 pm
GOL	s_7	10:58 pm	11:16 pm
AZUL	s_8	11:14 pm	11:24 pm

More information is needed: new schedule slot created in the process RBS of GDP/CDM. The schemes can be seen in Table 7.

To choose the best preference ordering over slots, each airline groups its own flights using results from $R_F(f)$ and decides the adjustment of the earliest possible arrival time e_f. The only restriction is operational constraints as all passengers must be present, crew is on duty and aircraft fueled and ready to go.

In this part of processing the adjustment e_f of each flight is compared to each slot and analyzed if allocation is possible. In a positive case, that slot is a candidate for allocation with that fight and is put in its ordering preference list, in a increasing order of slot time. This is done for all flights and all slots.

With the reference orderings of flights were calculated, the pre-processing algorithm processes the preference orderings of slots according to the airport's definitions in Eqs. 4 and 6.

The parameter c depending θ are used to adjust and normalize the delay is calculated by Eq. 5. This strategy allows the calculation of $R_S(f)$ through a standard delay it has a minimum value 1 (Table 8).

Table 8. Results high-priority order for airport

Flight ID	Slot schedule after RBS $st(s)$	Estimated slot time of arrival $at(f)$	(c)	$D_S(f)$	$R_S(f)$	H^*
f_8	11:16 pm	11:58 pm	10	1.8	8013	1
f_1	11:28 pm	10:28 pm	10	1	216	2
f_8	11:24 pm	11:14 pm	10	1	95	3
f_4	10:52 pm	10:46 pm	10	1	95	4
f_3	10:44 pm	10:35 pm	10	1	89	5
f_2	10:36 pm	10:32 pm	10	1	89	6
f_5	11:00 pm	10:49 pm	10	Canceled	Canceled	0
f_6	11:08 pm	10:55 pm	10	Canceled	Canceled	0

The delay adjustment is used to maximize the importance given to delayed flights over the amount of passengers and delays. Equation 3 calculates this information using passenger data capacity multiplied by the occupancy rate. Then, $R_S(f)$ is calculated by the high value to set delay resulting power. This mode $R_S(f)$ is responsible for enabling the airport set the order of their priority, their order of preference.

To calculate the preferences of the hourly airport slots use f information set and sort priority. Table 9 shows the result of this process.

Table 9. Preference ordering of slot $\succ s$

Flight ID	H^*	Adjusted e_f	Slot ID	Slot scheduled	$\succ s$
f_8	1	10:58 pm → 10:52 pm	s_1	10:28 pm	f_1
f_1	2	11:14 pm → 11:08 pm	s_2	10:36 pm	f_1, f_3, f_2
f_8	3	10:46 pm	s_3	10:44 pm	f_7, f_1, f_4, f_3, f_2
f_4	4	10:35 pm	s_4	10:52 pm	$f_7, f_6, f_1, f_4, f_3, f_2$
f_3	5	10:32 pm	s_5	11:00 pm	$f_7, f_6, f_1, f_8, f_4, f_3, f_2$
f_2	6	10:36 pm	s_6	11:08 pm	$f_7, f_6, f_1, f_8, f_4, f_3, f_2$
f_5	0	Canceled	s_7	11:16 pm	$f_7, f_6, f_1, f_8, f_4, f_3, f_2$
f_6	0	Canceled	s_8	11:24 pm	f_1, f_3, f_2

*High-priority Order for Airport

Allocation step is the main process of the proposed model. It uses the information provided from the sets of flights, slots, flight preferences, slot preferences, and original owner of slots.

We can see in Table 10 the initial scenario from allocation process. The allocation process starts with a proposal and acceptance step. First, f_1 proposes to his most preferred option, s_1. All other flights follow the same step, f_2 and f_3 propose to s_2, f_4 and f_7 propose to s_4, f_6 to s_5 and finally f_8 to s_6. As s_1 is empty, it accepts proposal and makes a pair with f_1. The slot s_2 prefers f_3 instead of f_2, and makes a pair with f_3 rejecting f_2, and so on. Table 11a shows the final allocation after the first round and Table 11b shows the result of the last round.

Table 10. Sets of flights preferences $\succ F$, slots preferences $\succ s$, and original owner of slots (O)

Airline	Airport	Owner
$P(f_1) = s_1 \succ s_2 \succ s_3 \succ s_4 \succ s_5 \succ s_6 \succ s_7 \succ s_8$	$P(s_1) = f_1$	$O(s_1) = TAP$
$P(f_2) = s_2 \succ s_3 \succ s_4 \succ s_5 \succ s_6 \succ s_7 \succ s_8$	$P(s_2) = f_1 \succ f_3 \succ f_2$	$O(s_2) = AZUL$
$P(f_3) = s_2 \succ s_3 \succ s_4 \succ s_5 \succ s_6 \succ s_7 \succ s_8$	$P(s_3) = f_1 \succ f_3 \succ f_2$	$O(s_3) = AZUL$
$P(f_4) = s_4 \succ s_5 \succ s_6 \succ s_7 \succ s_8$	$P(s_4) = f_7 \succ f_1 \succ f_4 \succ f_3 \succ f_2$	$O(s_4) = AZUL$
$P(f_7) = s_4 \succ s_5 \succ s_6 \succ s_7 \succ s_8$	$P(s_5) = f_7 \succ f_1 \succ f_4 \succ f_3 \succ f_2$	$O(s_5) = GOL$
$P(f_8) = s_6 \succ s_7 \succ s_8$	$P(s_6) = f_7 \succ f_1 \succ f_8 \succ f_4 \succ f_3 \succ f_2$	$O(s_6) = GOL$
	$P(s_7) = f_7 \succ f_1 \succ f_8 \succ f_4 \succ f_3 \succ f_2$	$O(s_7) = GOL$
	$P(s_6) = f_7 \succ f_1 \succ f_8 \succ f_4 \succ f_3 \succ f_2$	$O(s_8) = AZUL$

Table 11. DA-CDM process

(a) DA-CDM process first round

Slot	Flight	Owner	e_f
s_1	f_1	TAP	1
s_2	f_3	AZUL	2
s_3	Empty	AZUL	
s_4	f_7	GOL	4
s_5	Empty	GOL	
s_6	f_8	AZUL	6
s_7	Empty	AZUL	
s_8	Empty	GOL	

(b) DA-CDM process result

Slot	Flight	Owner	e_f
s_1	f_1	TAP	1
s_2	f_3	AZUL	2
s_3	f_2	AZUL	2
s_4	f_7	GOL	4
s_5	f_4	GOL	4
s_6	f_8	AZUL	6
s_7	empty	AZUL	
s_8	empty	GOL	

All slots were allocated respecting to the preferences of both airlines and airport managers. Due to the cancelation, the original sequence with the vacancies of a few slots that were maintained until the indicial of the execution of DA-CDM algorithm. If one airline did a manipulation of the cancellation, i.e. without true communication, other airlines may be prejudiced. In this case, *Compression* allows a more equitable distribution for different airlines. The available slots can be used by other airlines. This important property of DA-CDM may redistribute the final slot allocation of the flights.

By comparing the results of Tables 3 and 11, even the airline GOL has not been benefited by the non-communication of a flight cancellation (f_6), AZUL airline was also with a penalty by the flight f_4, which could not have been allocated to a previous slot in 16 min ahead. With the application of DA-CDM, as in Table 11, our approach corrected this manipulation and allocated the flight AZUL to take the slot of GOL.

6 Conclusion

Collaborative Decision Making is an important paradigm to improve the efficiency of airport traffic flow management. Based on this concept, this paper reformulated the DA-CDM model with the theory proof of the stable manipulation of the politic mechanism in compression step of slot allocation. With this extension, it allows the user to include preferences of airport managers in addition to the ATC units and airlines.

The theory demonstration and case study approved that despite the DA-CDM being a stable process is subject to manipulation by its agents through the non-disclosure of information, or even untrue information that may lead to differences in preferences lists of market players, thus causing a better matching or even harming other agents.

Future work will modify the DA-CDM model in such a way that it can include more functions and performance measures for airport managers in order to make it unmanageable. Moreover, analysis time and complexity should be with performed DA-CDM approach that can be modified to provide the best results through Pareto efficiency. Finally, the aircraft allocation effects can be analyzed under different scenarios and type of manipulation that may incur in the market can be studied and used taking advantage of real-world data.

References

1. Abdulkadiroğlu, A., Pathak, P., Roth, A.E.: The New York City High School Match. In: American Economic Review, Papers and Proceedings (2005)
2. Arruda, J.A.C., Weigang, L., Barros, A.: Fairness analysis with flight cost impact using reinforcement learning approach. J. Braz. Air Transp. Res. Soc. **8**, 9–27 (2012)
3. Arruda, A.C., Weigang, L., Milea, V.: A new airport collaborative decision making algorithm based on deferred acceptance in a two-sided market. Expert Syst. Appl. **42**(7), 3539–3550 (2015)
4. Aumann, R.J.: War and peace. Prize lecture, 8 December 2005
5. Ball, M., Hoffman, R., Knorr, D., Wetherly, J., Wambsganss, M.: Assessing the benefits of collaborative decision making in air traffic management. In: Donohue, G., Zellweger, A. (eds.) Air Transportation Systems Engineering, pp. 239–250. American Institute of Aeronautics and Astronautics Inc, Reston (2001)
6. Brinton, C., Provan, C., Lent, S., Prevost, T., Passmore, S.: Collaborative departure queue management. In: Ninth USA Europe Air Traffic Management Research and Development Seminar, pp. 1–12 (2011)
7. Crespo, A.M.F., Weigang, L., Barros, A.G.: Reinforcement learning agents to tactical air traffic flow management. Int. J. Aviat. Manage. **1**(6), 145–161 (2012)
8. Cruciol, L., Arruda, J., Weigang, L., Leihong, L., Crespo, A.: Reward functions for learning to control in air traffic flow management. Transp. Res. Part C: Emerg. Technol. **35**, 141–155 (2013)
9. Fiani, R.: Teoria dos Jogos: com Aplicações em Economia, Administração e Ciências Sociais, 3 Edição. Elsevier, Editora Campus, Rio de Janeiro (2009)
10. Gale, D., Shapley, L.S.: College admissions and the stability of marriage. Am. Math. Mon. **69**(1), 9–15 (1962)
11. Hoffman, R.L.: Integer programming models for ground-holding in air traffic flow management. Ph.D. thesis, Department of Management Sciences, University of Maryland, College Park (1997)
12. Kesten, O.: On two kinds of manipulation for school choice problems. Econ. Theory **51**, 677–693 (2012)
13. NOBEL: Stable matching: theory, evidence, and practical design. The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2012, Roth, A.E., Shapley, L.S.: Advanced Information (2012). www.nobelprize.org/nobelprizes/economic-sciences/laureates/2012/popular-economicsciences2012.pdf
14. Roth, A.E.: The economist as engineer: game theory, experimentation and computation as tools for design economics. Econometrica **70**(4), 1341–1378 (2002)
15. Schummer, J., Vohra, R.V.: Assignment of arrival SLOTS. Am. Econ. J. Microecon. **5**(2), 164–185 (2013)
16. Sönmez, T., Utku Ünver, M.: Matching, allocation, and exchange of discrete resources. In: Benhabib, J., Bisin, A., Jackson, M. (eds.) Handbook of Social Economics, vol. 1A, pp. 781–852. North-Holland, Amsterdam (2011)
17. Vossen, T., Ball, M.: Optimization and mediated bartering models for ground delay programs. Nav. Res. Logist. **53**(1), 75–90 (2006)

Technology Monitoring System to Improve the Competitiveness and Innovation in Colombian Companies

Alvaro Espinel Ortega[1(✉)], Adriana Marcela Vega Escobar[1], and Víctor Martín García[2]

[1] Universidad Distrital Francisco José de Caldas, Bogotá, Colombia
{aespinel, avegae}@udistrital.edu.co
[2] Universidad Pontifica de Salamanca, Salamanca, Spain
victor.martin@upsam.es

Abstract. The way to boost the development of the regions has to do with making companies more competitive, in such a way that they can have access to the global market, this requires improving the traditional production processes, that's why it is a must to find new productive sectors in consistent with changes in the dynamics of global demand for goods and services. The aim of this paper is to show the first steps to give formal processes supported by research which are supported by science, technology and innovation.

This research focuses on the companies that can offer goods and services classified as micro, small and medium-sized companies (also known by the acronym "MiPymes" [1]), since these are the ones that have better chances of hitting directly to the society.

These companies require an accompaniment on the part of the state, in their different orders, so that their competitiveness can be improved on the basis of innovation, therefore, it is suggested that the process of observation of the environment will be followed by experts and with the help of a computational tool developed for this purpose called "technological monitoring system, multi-application". Reference [2], which makes it easy to do the diagnosis of the current status of the companies based on the measurement of the technological gap, this process is oriented to discover potential sectors, to modernize the traditional or reject those that lose their validity, in order to improve the quality of the products or services offered, improve the efficiency of production processes and increase the competitiveness of companies.

Although it is true there are many software tools for technology monitoring processes, this is a system that has the ability to offer a free service to small businesses that do not have resources to pay for such studies. It is intended that the system is used by State institutions, as a free public service to support "MiPymes" [1] as part of a Public Policy.

Keywords: Business process management · Technology monitoring · Competitive and business intelligence · Innovative business models

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 312–328, 2015.
DOI: 10.1007/978-3-319-21009-4_24

1 Introduction

The Colombian companies either are oriented to the production of goods or services, especially those classified as micro, small and medium-sized enterprises, which are usually referred to as the acronym "MiPymes" [1], they do not have friendly computational tools to perform follow-up processes to different aspects related to technological advances, which have to do with its production facilities, this is the reason why many of these companies have currently obsolete technologies or even they do not make use of any of them.

With such an observation, conducted on the basis of computational tools, the Colombian state and the business could have benchmarks, in order to propose policies to support those sectors that are in crisis as well as discover new sectors potentially important for supporting and reject those that lose their validity.

This observation is proposed to be carried out using a "Technological Monitoring System, multi-application" [2], in order to measure the technology gap in the Colombian companies, that is to say, that it can be configured in such a way that applies to any kind of company, either for goods or services, regardless of the economic activity that you perform and the productive chain to which it belongs.

Learn how are companies in the country, compared to the companies of a similar nature in the international context, through the selection of a set of indicators based on international relations, in order to measure the technology gap, could allow us to make decisions in order to establish policies on the part of employers and state policies to decrease this gap and increase the competitiveness of enterprises through better use of human and economic resources. To measure the gap it will be taken as a reference the developed countries in the Americas and Europe, Asian countries just like Japan and China and emerging countries in Latin America and Mexico.

2 Technological Monitoring: Conceptual Revision

2.1 Competitiveness and Innovation

The best indicator of the competitiveness of a country is the per capita income of its inhabitants. "The competitiveness of a nation is its ability to produce goods and services in international markets, maintaining or increasing real incomes to their citizens. Competitiveness is the basis of the standard of living of a country", OECD (Organization for Economic Cooperation and Development) [3]. In development of this research one of the authors interviewed in Bogotá, Colombia with two important people in this organization: Fred Gault and Dominique Guellec, former directors OECD [3] from the United Nations University. These two people participated in the Forum: Connecting to Colombia from innovation, conducted in September 2011. Competitiveness means a single value offered to consumers, which lets you raise at the same time the income level of workers. This means: higher wages for workers, higher profits for the producers, greater level of investment in both capital and in training and technology, and higher levels of production and exports, that is competitiveness. A country is called itself "competitive" due to its companies [4].

The competitive development involves two processes:

Analytical Process: Strategic Design, technological monitoring, market research, product positioning and distribution.

Operating Process: Group cohesion, common agenda for action. The minimum requirements to achieve competitiveness are: political feasibility, financial viability, social inclusion and socio-environmental respect. In the final stage of the last century, the industrial innovation and R + D + I (Research + Development + Innovation) show a step forward toward a generation based on knowledge, learning and information flows, which closes even more the relationship between the company and the world around it.

2.2 Data, Information, Knowledge, Technology and Innovation

In order to perform a brief description of the conceptual part of the terminology used to describe the knowledge from the data and its relationship to innovation, the definitions (Ponjuan, 1998), related to these aspects are quoted as they are essential to be taken into account when a person studies in the field of technological monitoring [5, 6].

Data: Quantitative or qualitative characteristics that are obtained from the source of information. They are analyzed to study in detail the characteristics of the source.

Information: Related data, processed and interpreted on a particular or specific subject.

Knowledge: It is the result of processing the information, which is generated at the inner person that processes it. It is a set of information developed on the basis of experience and applied to an action, it allows the decision-making.

Knowledge Management: It is defined as the process of continually manage knowledge of all kinds to meet present and future needs, identify and exploit knowledge resources both existing and acquired to develop new opportunities [7].

Technology: It is the structured knowledge that is applied to get results. It can be said that it is the practical application of the sciences.

Innovation: It is the result of the introduction economically and socially useful of the new knowledge. Certain authors, characterize this generation by the integration of different components of the industry (science, technology, market and society), flexibility of the processes and above all, efficient management of the information that allows the entity to have a strategic knowledge from the data obtained from the computing management. Carter [7], argues that innovation is a process of accumulation of knowledge ("Knowhow") as well as learning by providing organizations obtain great benefits from the rapid information management, where the connection between the internal of the company and the outside (suppliers, distributors, customers), is separated into parallel and real-time [7].

Based on the identification, acquisition, classification and maintenance of knowledge, it must be taken into account its use to achieve the goals of the organization and to find competitive advantages [8].

"The current company is defined as a set of tangible and intangible assets, where the latter are increasingly having more importance and effectiveness in creating value for the company; intangible assets that are the result of the incorporation of knowledge, intellect, to the various productive activities of the organization" [9] (see Fig. 1).

2.3 What Aspects Must Be Monitored?

There are several aspects that must be monitored:

Technological Aspects: The scientific and technological advances, the result of the basic and applied research, the products and services, manufacturing processes, materials, the processing chain, the technologies and information systems.

The company that operates in an international framework must be aware of what and who they are working with, the people of technical centers, universities and laboratories related to the area of its competence [10]. In the analysis of alternative products and technologies, the person who observes should know the features that match its products, the processes of transformation that it suffers, its design, the range of related products and/or complementary, etc. The investment in I + D of their competitors and companies of the value chain, the number of scientists and engineers, scientific and technical publications, and its patents [11].

Competitive Aspects: Analysis and follow-up of current and potential competitors. The fate of their investments, their products, distribution channels, response times, type of clients and degree of satisfaction, its organization, its financial ability, etc. The sector value chain and the situation of the company and its strength in the value chain.

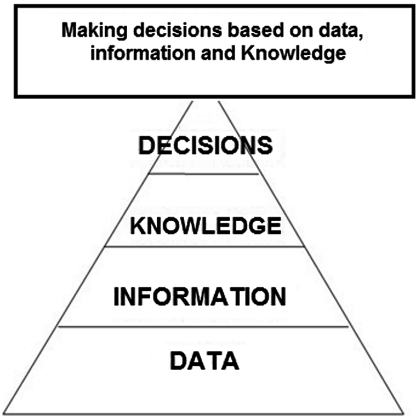


Fig. 1. Informational pyramid. Source: [5]

Commercial Aspects: Markets; customers, the evolution of their needs, their solvency; suppliers, their strategy to launch new products, their suppliers; labor in the sector and in the value chain.

Aspects of Environment: The laws and regulations, non-tariff barriers; the environment and the care evolution; culture: behind any decision there are people; politics, sociology, etc.

2.4 The Technological Monitoring and Competitive Intelligence

The word competitive intelligence tends to replace the term technological monitoring and therefore it constitutes a more active aspect, it presents a more elaborate information and better prepared for the decision-making. Between the two disciplines there is a difference in tone, while the technological monitoring puts the emphasis on the search and obtaining relevant information for decision-making, the competitive intelligence refers to the same process, but with the emphasis on decision-making. In this sense, the monitoring should help to identify, in advance, the type of technological innovation that can affect productivity.

Differences Between Technological Monitoring and Competitive Intelligence. The Technological Monitoring can be understood as the process of observation, search, identification, analysis, processing and distribution of the information.

The Competitive Intelligence must be understood as the art of applying the knowledge acquired from the Technological Monitoring in obtaining competitive advantages to lead the market.

The monitoring technology and the competitive intelligence must be two permanent and continuous processes, not sporadic and they must be performed by all members of the organization. The Competitive Intelligence you should start from people, to be integrated in the working groups, the tactical and operational departments, primarily in the areas of leadership. Competitive Intelligence is a set of strategies.

2.5 The Process of Technological Monitoring and Competitive Intelligence

The dynamics and the saturation of markets are leading to the failure of many organizations that do not have methodologies and tools of analysis from their environment, which means that perform processes of technological monitoring and processes of competitive intelligence aren't applied (see Fig. 2).

3 Problems to Innovate and Improve Competitiveness

In general for the countries of Latin America and the Caribbean, among others the following problems have been identified, when performing processes of innovation in order to improve competitiveness:

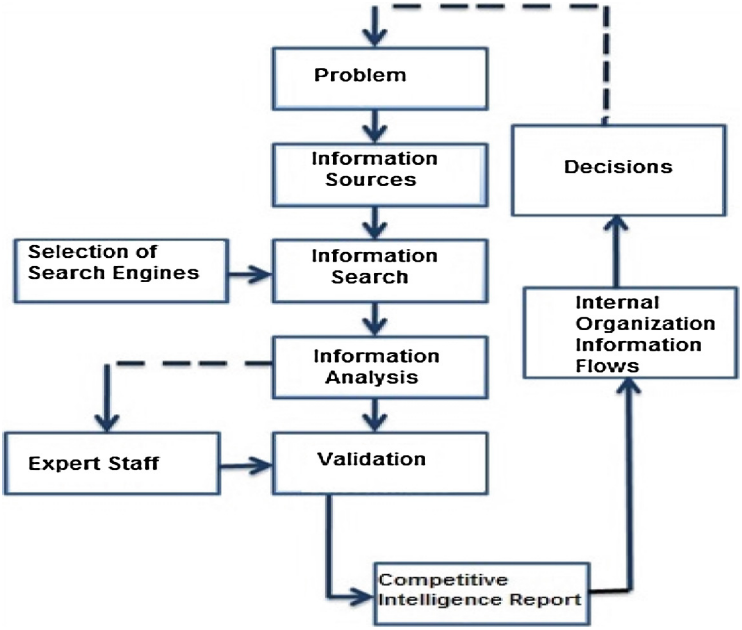


Fig. 2. The process of technological monitoring and competitive intelligence. Source: [12]

At the national or international level it is not a standard that establishes what indicators must be considered in determining the technological gap, understood as the degree of lag or advancement that exists between the world's leading companies and firms that are pretended to be studied. On the other hand, there are not scales to quantify the extent of the degree of innovation, in companies, products or services.

They are virtually non-existent State policies relating to motivate the realization of technological monitoring exercises or at least observations in the fields of science, technology and innovation that will enable to improve their processes.

Studies leave aside methodological aspects of human resources, being especially complicated for small businesses not possess the resources to be able to pay the conduct of these studies that is why they are at a disadvantage compared to its competitors. It is here where the State should lead this process in order to support the micro, small and medium-sized enterprises known as "MiPymes" [1], because they are a great generator of employment in the country.

4 Study of Technological Monitoring and Gap Measurement

In the system proposed the study of Technological Monitoring has three components: the employer, a coordinator and an expert, they are responsible for interacting with the system (see Fig. 3):

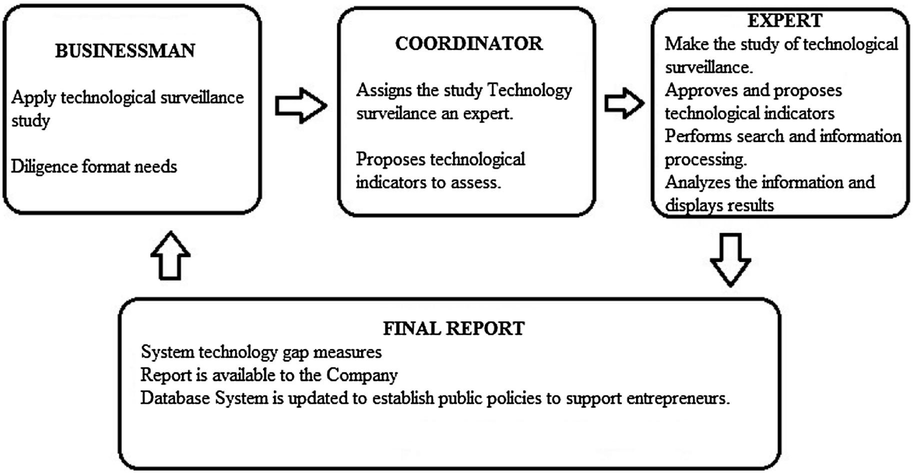


Fig. 3. Generation process of the study of Technological Monitoring. Source: Authors

Measurement of the Technological Gap. The realization of the measurement of the technological gap is based on the indicators, for this purpose the system is designed in such a way that they can be created in a free manner from the Coordinator of the system, but in accordance with the considerations of an expert or an entrepreneur. The system allows the generation and the calculation of indicators through comparisons and connections in order to display graphically the technological gap between the companies in the city of Bogotá and the rest of the world.

Technological Indicators. The system indicators are originated in the need to understand the behavior of enterprises in relation to the activities of innovation; therefore they are related to innovation in product, process, management and marketing. They pretend to establish units of measurement that will guide the design of strategies to improve the performance of the companies.

Process Indicators of Management. The following chart summarizes the process that should take place within the system developed for the management of technology indicators, for further comparison and measurement of the technological gap (see Fig. 4).

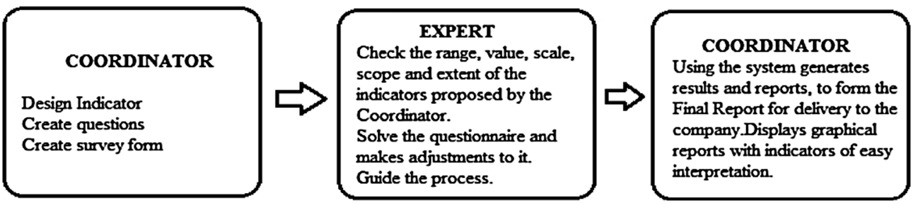


Fig. 4. Management process or administration of indicators. Source: Authors

Table 1. Description of the types of indicators that handles the system. Source: Authors

Field	Description
Name	Name of indicator.
Date Created	Date indicator is created.
Type of indicator	Indicates that sector or business area focuses indicator. The types of indicator are dynamic and can be created according to the needs of the system. Types of indicators are: Product, Process, Marketing, Organizational Management.
Scope Indicator	Indicates the field on which the indicator is calculated: Single, town, city, ISIC, national, international, global, etc.
Measure Indicator	Indicates the measurement of indicator: Qualitative, quantitative.
Range Indicator	Percentage, magnitude, scale, index.

The indicators are the benchmarks on which you can perform quantitative or qualitative measurements. In case of being quantitative scales, numerical values can be assigned to perform their respective allocations.

Design of the Indicator. The design of the indicator is a task of the coordinator that can be assisted by experts, entrepreneurs and other coordinators. This may be the most important task for the measurement of the technological gap, because it determines which and how you intend to measure and compare (see Table 1).

5 Model System for the Proposed Technological Monitoring

In accordance with the requirements for the companies found "MiPymes" and following a process for design and software development methodology based on the RUP® [13], a system for the monitoring of technology with the following subsystems was designed:

Management subsystem "MiPymes", subsystem of user management, document management subsystem, subsystem of indicators, measurement subsystem of Technological Gap.

For this article, an analysis of the system will be presented from the perspective of the generation of a study of technological monitoring, to measure the gap.

5.1 Use Case Diagram

5.1.1 Use Cases Diagram Overview of Technological Monitoring System (see Fig. 5)

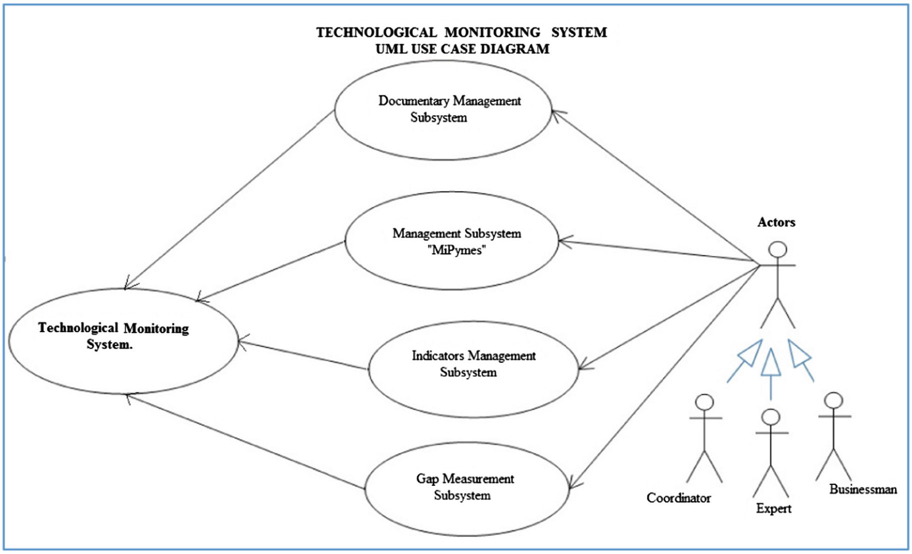


Fig. 5. Use case diagram for technological monitoring system. Source: Authors

5.2 Implementation of the Prototype

The Technological Monitoring System was implemented in a multi-level architecture (3 basic levels) with independence between the presentation or graphical user interface (GUI), which in this case is a graphical user interface web, which can be run with any web browser or "browser", on both PC and mobile devices, application logic or business logic (processes) that was implemented with C# programming language and persistence or storage of data, which was implemented with Microsoft® SQL SERVER [14], using the Visual Studio NET [15] (see Fig. 6).

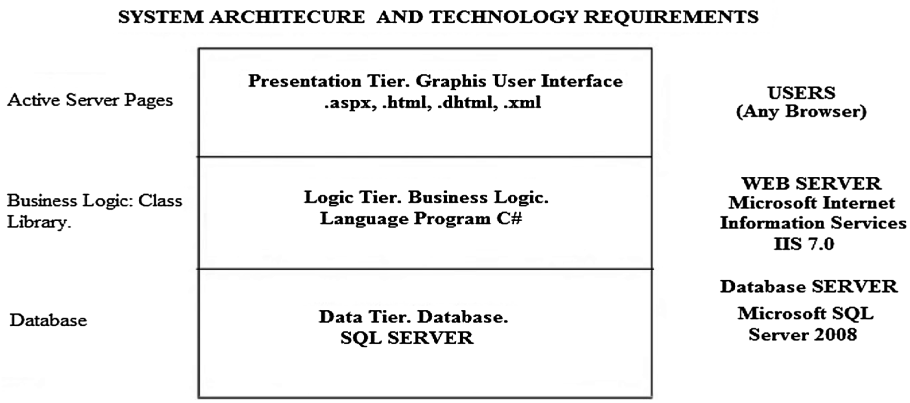


Fig. 6. System architecture implemented. Source: Authors

5.3 Graphical Model of the User Interface (GUI)

The GUI is displayed with the system in operation in Fig. 7.

Fig. 7. Model graphic user interface. The GUI is in Spanish. Source: Screen capture system operating. Author

5.4 Subsystem of Measure of the Gap

The system was designed in such a way that the gap can be measured based on the comparison of indicators.

In the graph bar it can be seen the different indicators for the company in study (e) called as (Yen), which can be obtained from others from the survey that the employer completes in the system and in which the process has been previously illustrated. In the case of the indicators of the country of reference (p), they are called (Ypn). It can be seen from the above chart that the comparison is not only limited to technological indicators but of any kind. Visually in the system the gap can be observed, but for the purposes of better understanding especially for the employer, it is performed a representation that includes the measure which can be seen conceptually in the following figure (see Fig. 8):

5.5 Measure of the Magnitude of the Gap and Interpretation of Results in Accordance with the Scope of the Indicator

If the scope of the indicator is taken into account, it can be seen that this system not only allows you to measure the technology gap, but also depending on the type of indicator it could measure other aspects such as: human resources, degree of innovation, physical resources, infrastructure, and in general any indicator that can be measured quantitatively or qualitatively.

In accordance with the comparison of indicators it can be seen in the following figure, the manner in which the magnitude of the gap through the difference of indicators is established, which, in the case of the indicator **n-**esimo, it can be determined

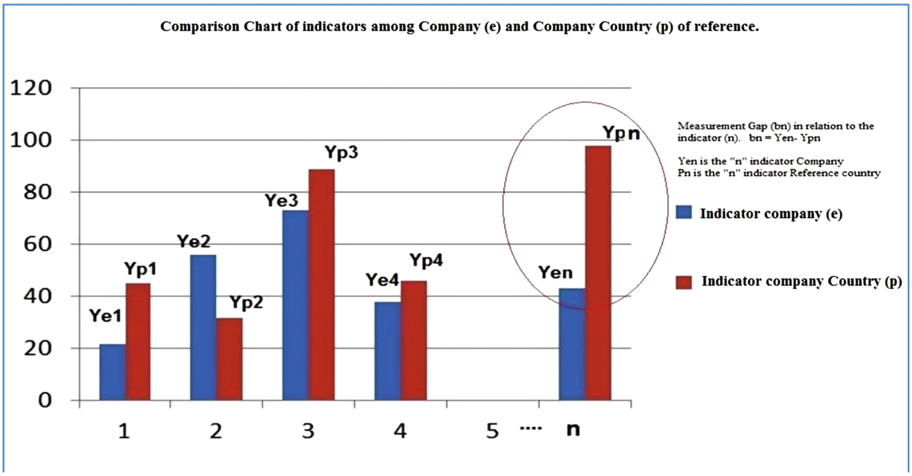


Fig. 8. Comparison chart of indicators among company (e) and country (p) of reference. Source: Authors

as: **bn = Yen-Ypn**, where **bn** is the magnitude of the gap for the indicator **n**-esimo, expressed in units of the indicator, **Yen**, is the magnitude of the indicator **n**-esimo for the company in study **(e), Ypn**, is the magnitude of the indicator **n**-esimo for the country (p) (see Fig. 9).

It can be seen in the chart how the magnitude of the gap for each of the indicators that are evaluated is established. For the interpretation of results the following considerations are taken into account:

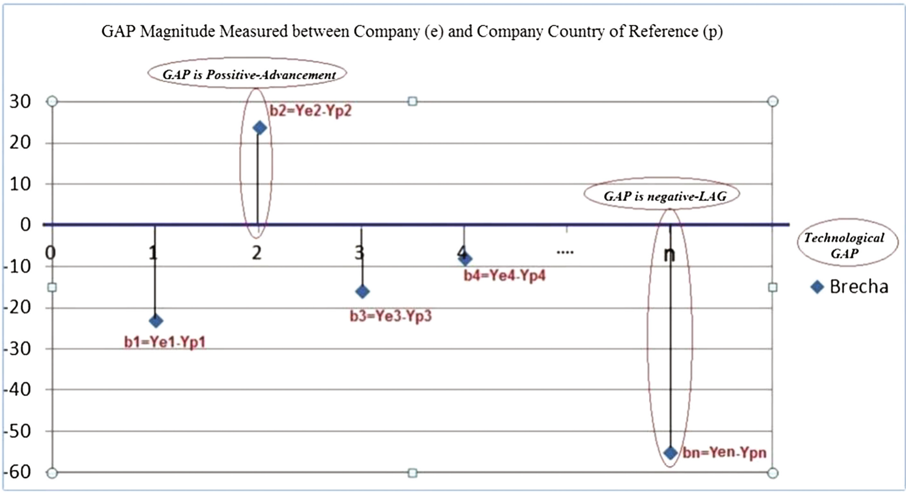


Fig. 9. Gap measurement (b), between enterprise (e) and country (p) of reference. Source: Authors

Table 2. Table proposal for classification of businesses in accordance with the extent of the gap. Source: Authors

Result Gap	Proposed Name (−)	Proposed Name (+)
(Measurement)	Lag	Advanced
(0–20]	Less Lag	Minimum Advancement
(21–40]	Moderate Lag	Moderate Advancement
(41–60]	Severe Lag	Intermediate Advancement
(61–80]	Critical Lag	Best Advancement
(81–100]	Extreme Lag	Great Advancement

Possible Conditions:

bn > 0: Gap is positive (+). Company or a country study is in better position as the reference.

Status: Advancement.

bn < 0: Gap is negative (−). Company or a country study is in worse condition than the reference.

Status: Lag.

bn = 0: Gap is zero (0) Company or country is in a position similar to the reference. **Status: Similar.**

With the expert support tables with ranges may be assigned and on the basis of this be able to make decisions.

The previous classification will allow entrepreneurs have a clearer idea of its position in relation to companies or countries which they compare because the relation of percentages and the name proposed for the classification provides reports with greater precision of the status of the company in relation to the extent of the gap for each of the indicators in study (see Table 2).

6 Test of the Prototype

Tests and validation of the prototype were conducted with the support of the Secretary of Economic Development of Bogotá D.C. For the application of the tests the collaboration of more than fifty (50) companies was helpful, which filled out a survey based on the indicators selected for performing the process of measurement of the gap. Sample Data:

Countries that serve as a reference for testing: Republic of Colombia, Spain, Korea, USA.

Population: 22.172 Registered Companies "MiPymes"

Sample Size: 50 "MiPymes"

Questionnaire Questions: 37

Indicators used: 31

$$TmAjustado = \frac{Tm}{1 + \frac{Tm-1}{Poblacion}} \qquad c = \sqrt{Z^2 * p * (1-p)/Tm}$$

c = Confidence Interval in decimal. (Example c = 0.04 = ± 4 %)
Z = Z (1.96 for 95 % Confidence Level)
p = Percentage of success or not in responses, taking as 0.5 for this case, where it is completely random.
Tm = Sample Size
TmAjustado = Sample Size Adjusted

With the above data the confidence interval is: 0.138592929 = 13.85 %

7 Measurement Results of the Technological Gap

7.1 Test Measure Gap Colombia – Spain for Three (3) Basic Indicators

Below are the results of tests for measuring the gap between Colombia and Spain, for three basic fundamental indicators that are highly related to the technology gap.

These three (3) indicators that appear at first sight to be very basic, are the true essence which indicates how entrepreneurs can use Internet taking into account the two (2) countries under study. The indicators evaluated are: use of computers, internet use and Web presence of the companies (see Fig. 10).

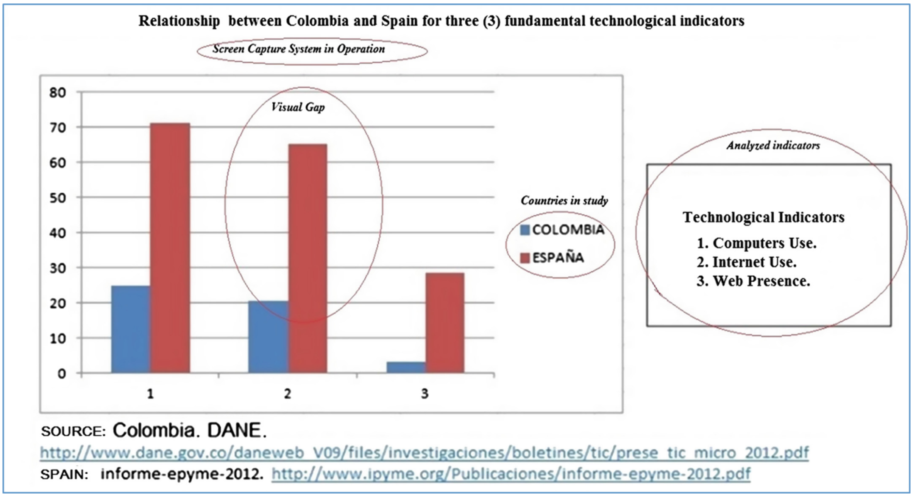


Fig. 10. Relationship between Colombia and Spain for three (3) fundamental technological indicators. Source: Authors. Screen capture system operating

In the following graph it can be seen the magnitude of the gap and the interpretation of the results (see Fig. 11).

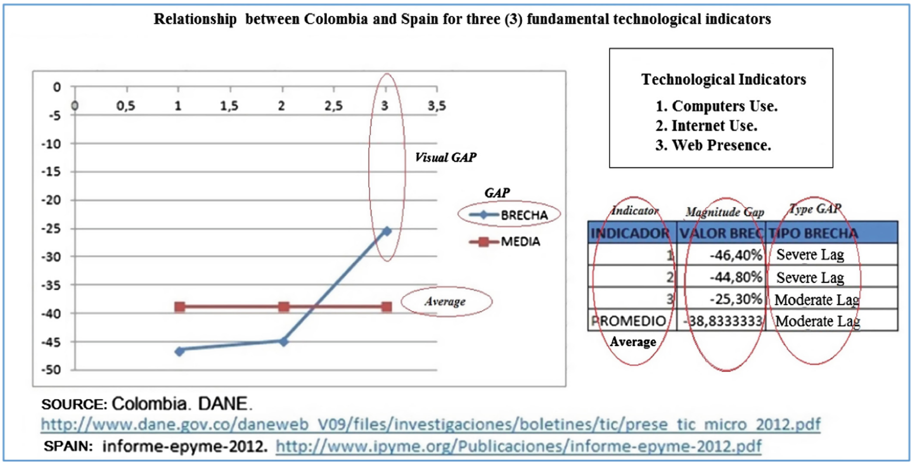


Fig. 11. Measure of the gap and interpretation of results for three (3) technological indicators. Source: Authors. Screen capture system operating

From the previous graphs it can be concluded that fundamentally in relation to Spain, Colombia is lagging behind in a high degree. The analyzes and comparisons with other countries must be performed by a team of experts who must be involved in the Technological Monitoring System and must include policies of State.

7.2 Test with a (1) Single Indicator and Four (4) Countries

In order to test the system with data from institutions such as the World Economic Forum (World Economic Forum-WEF), for one (1) single indicator in this case the

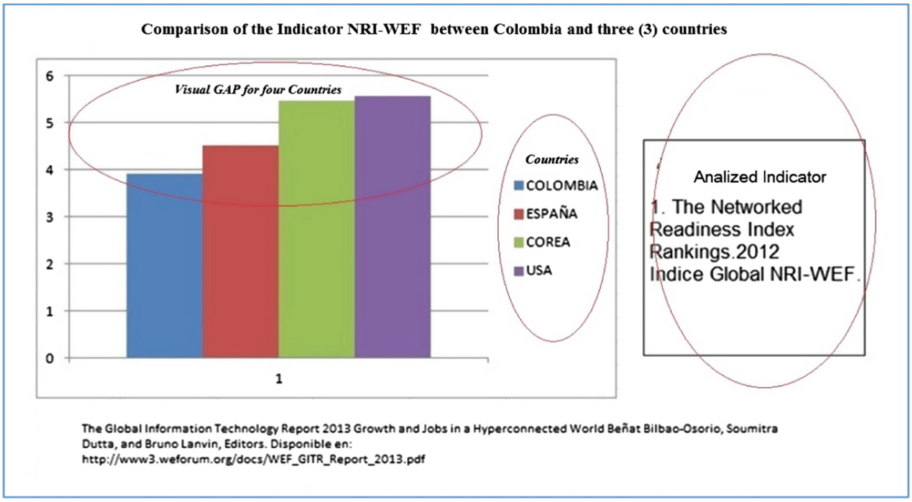


Fig. 12. Comparison of the indicator NRI-WEF between Colombia and three (3) countries. Source: Authors. Screen capture system operating.

Networked Readiness Index (NRI, which measures how the country rely on the Internet and ICT to improve their competitiveness, it is observed in the previous graph (see Fig. 12).

The following graph shows the extent of the gap: (see Fig. 13).

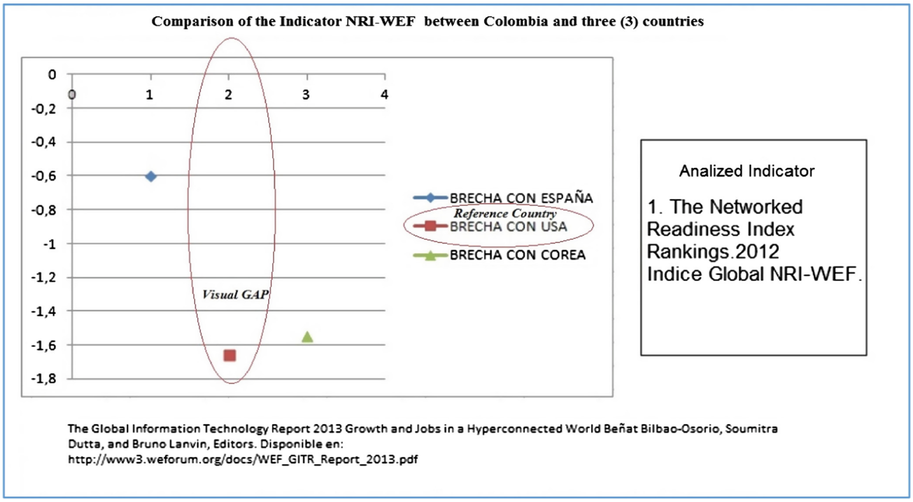


Fig. 13. Measure of the gap of the Indicator NRI-WEF between Colombia and three (3) countries. Source: Authors. Screen capture system operating.

In this analysis of results it is observed that Colombia is in lag, so the experts and the Colombian State are responsible for making relevant decisions in order to reduce the gap.

8 Conclusions

It was designed, implemented and tested a technological monitoring system based on indicators, which can be handled, it means, create, consult, modify, delete, in accordance with the guidelines of a expert that guides the studies. These indicators do not have restrictions of any kind and can be created freely in accordance with the convenience and are related to human resources, technology, science, innovation and of course with economic and financial indicators.

Taking all of the above in consideration not only to make measurement studies of gap in technological aspects, but other issues such as human resource. For example, a small company with fewer than 20 employees in the field of manufacturing of garments can be measured, also it can be determined how many professionals, specialists, doctors, etc. it has, to compare the national context with the international one and observe how this affects productivity.

In order to take in what it was said above, the system allows you to parameterize each of the indicators and to establish whether they are qualitative or quantitative, in the latter case it allows you to configure the respective scales in order to be able to make the measurement.

In tests applied of the prototype with different countries, there is a technological backwardness, which validates what is perceived by the community in general. This work aims to raise awareness to facilitate the creation of State policies and create awareness in the employer of the importance of technological support to observe the environment. Thus, it is important that in addition to meet the particular needs of each employer in the topic of technological monitoring, there can be consolidated information that would allow those who handle the policies related to each of the fields investigated and take the actions to perform a continuous improvement.

The result of the measurement of the gap between Colombia and Spain for three (3) fundamental indicators: (1) Use of computers, (2) Use of Internet and (3) Web Presence, Colombia goes wrong evaluated, what which should suggest the Government to implement policies to promote the use of information technology, education campaigns, facilities for equipment procurement, in order to reduce this gap because otherwise the country has no chance of competing or innovate and its population will be at a disadvantage with other countries.

References

1. Ministerio de Comercio, Industria y Turismo, República de Colombia., Mipymes, Portal Empresarial Colombiano, Ministerio de Comercio, Industria y Turismo, República de Colombia, 01 January 2011. http://www.mipymes.gov.co/. (Accessed 19 June 2014)
2. Espinel Ortega, A.: Modelo de Sistema para la Vigilancia Tecnológica de Aplicación Múltiple, Madrid (España), Universidad Pontificia de Salamanca (2014)
3. OCDE, Manual de Oslo (2009). http://www.oecd.org/science/inno/2367580.pdf. (Accessed 23 January 2013)
4. Cabanelas, J.: La dirección de empresas en un entorno abierto, dinámico y complejo (1997). José Cabanelas Omil. (Accessed 16 April 2012)
5. Ponjuan, G.: Gestión de Información en las Organizaciones, Editorial Pirámide (1998)
6. Porter, M.E.: Estrategia competitiva. Técnicas para el análisis de los sectores industriales y de la competencia, México: CECSA (1980)
7. Carter, L.J.: Benchmarking and Analysis of Science and Technology. In: SCIP CTI Symposium (1997)
8. Porter, M.: Ventaja competitiva. Creación y sostenimiento de un desempeño superior, México: CECSA (1985)
9. Porter, M.: The Competitive Advantage of Nations. Mac Millan Press, Londres (1990)
10. Polanco, Y.: Innovación en las organizaciones, Noviembre 2010. http://www.slideshare.net/yumarys/creatividad-e-innovacion-en-las-organizaciones
11. Porter, M.: What is strategy? (1996)
12. Madrid. http://www.madrimasd.org/revista/revista17/tribuna/tribuna1.asp?imprimir=si (Último acceso: Noviembre 2010)

13. IBM, IBM Corporation, 10 December 2010. https://www.ibm.com/developerworks/rational/library/content/03July/1000/1251/1251_bestpractices_TP026B.pdf. (Accessed 15 December 2012)
14. Microsoft Corporation, Motor de Base de datos SQL Server, USA (2008)
15. Microsoft Corporation, Visual Studio 2008, USA (2008)

Modelling Knowledge Spillover Effects Using Moderated and Mediation Analysis – The Case of Czech High-Tech Industries

Jan Stejskal[1] and Petr Hajek[2(✉)]

[1] Institute of Economic Sciences, University of Pardubice, Studentska 95, CZ53210 Pardubice, Czech Republic
jan.stejskal@upce.cz

[2] Faculty of Economics and Administration, Institute of System Engineering and Informatics, University of Pardubice, Studentska 95, CZ53210 Pardubice, Czech Republic
petr.hajek@upce.cz

Abstract. This paper presents a new perspective on the evaluation of the contribution of high-tech industries. Standard macroeconomic indicators are not used in the evaluation, but rather the influence and importance of high-tech industries are analysed with the help of their spillover effects and the ability to create innovations. These are difficult to measure effects which significantly affect other businesses. Likewise, these high-tech industries are affected by the very businesses they work with. This paper presents evidence that if high-tech industries do not cooperate with other businesses and acquire their information only from internal sources, the result will be the creation of the smallest spillover effects. In the event, collaboration with universities and research institutions creates the greatest indirect effects. Furthermore, it was found that the most direct influence on innovative activities is information obtained on current markets. The paper presents evidence that drawing government aid or aid from EU funds has no effect on the creation of knowledge spillover effects in high-tech industries.

Keywords: Knowledge spillover · Knowledge process · Collaboration · Innovation · Structural equation modelling · Moderated mediation analysis

1 Introduction

The current economy of every country must, as part of globalising tendencies, face various changes, which come in a variety of evolutionary waves [1–3]. These changes also affect sources of competitive advantage or methods of achieving business goals or goals set by public policies directed towards maximising the welfare of the countries' citizens [4–7]. This has resulted in a gradual change in tools used to achieve the economic development of both businesses and regions or entire countries [8].

The first wave of evolutionary changes was aimed at attracting new investment (e.g., in Central Europe this is dated to the 1990s; in Western Europe it began decades earlier). The source of this development was massive investments in infrastructure and

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 329–341, 2015.
DOI: 10.1007/978-3-319-21009-4_25

various hard business elements [9]. It supported the building of a favourable business environment for incoming investors by minimising the initial costs of entry into a certain economy and beginning production within a very short time. This resulted in the creation of new jobs, achieving a certain desired performance. All this was massively supported by subsidies, subsidised loans and various tax breaks - thus public investment in private equity. Attention was not paid to the quality of the output, but rather to quantity and speed of processing [10, 11].

The second wave of evolutionary development builds on the previous stage and tries to avoid their mistakes and shortcomings. Again, we are dealing here with an important role played by the public sector, particularly when the local government supports specific business activities and projects, or establishes a specialised infrastructure for the expansion of existing production capacities [12]. There is also a change in orientation of the source of regional competitive advantage, which is quality products processed using the latest high-technology [13–15].

The third wave of development is focused exclusively on businesses with growth potential which are focused on high-tech production. This type of production is automatically associated with the creation and commercialisation of innovations by achieving a significant competitive advantage and realising economic benefits from the activities concerned [16]. The third stage of development occurs when it is economically very difficult for individual businesses to acquire the latest equipment and technology, to create a fundamental productive or process innovation and obtain a maximum economic benefit [17]. Therefore, at this stage, other development tools are used which are based on collaboration, using science and research to spur innovation, and creating industrial clusters and other forms of collaborative chains (confirmed by studies [18–20]). Emphasis is also placed on the use of spillover effects resulting from cooperation, social workers' capital, knowledge, and the ability to use tacit knowledge to enhance their own competitive advantage [21–23].

Recently, researchers have shown an increased interest in examining the effect of collaboration (both internal and external) on innovation activity [24]. A positive effect has been observed mainly due to the fact that collaboration boosts knowledge flows and knowledge spillovers [25–28]. Thus collaboration plays an important role in mediating the effect of knowledge acquisition on innovation activity. This has been demonstrated extensively for both manufacturing [27–30] and service sectors [31].

Here, we focus on a specific sector of medium-tech and high-tech manufacturing industries that have seen increasingly rapid growth in the last decade. This paper demonstrates that inter-organisational collaboration of these companies leads to a more efficient use of knowledge. This mode of collaboration also seems to be more important than collaboration with universities and research institutes. We analyse the effect of knowledge acquisition from various external sources and show, that under the influence of collaboration, side-effects arise - spillovers that facilitate innovative activity. With the aid of moderated and mediation analysis we can calculate them. This allows us to assess the importance and effectiveness of collaboration between companies in creating innovations such as with universities, R&D organisations or other groups of companies. It has been shown that if companies collaborate with universities and R&D organisations, greater spillover effects will be created. Furthermore, it was found that the greatest direct influence on innovative activity is obtained information on the current market.

This is done by examining both direct and indirect (mediating) effects (via collaboration) on innovation activity. Additionally, we test the moderating effects of both financial constraints and the receipt of public support which may be positively related to undertaking collaborative innovation.

The remainder of this paper has been divided into four sections. The paper first gives a brief overview of the evidence for knowledge spillovers in medium-high- and high-tech industries. The next section lays out the theoretical foundations of structural equation modelling, and a theoretical model for knowledge spillover analysis is proposed. Section 4 provides the results of experiments and Sect. 5 concludes this paper and discusses its implications.

2 Knowledge Spillovers in High-Tech Industries

Investment in the usage of knowledge, creation of new technologies, and the subsequent creation of innovations have been carried out in various industrial and service sectors for many years. They are not used exclusively in the so-called traditional businesses, where they were used for the first time during the 18th century in the textile industry in the UK [32]. During the second industrial revolution in Germany and the U.S., during the 19th and early 20th centuries, there was a significant use of new technologies for the production of innovations in the chemical industry and its use in dying fabrics [33]. In these areas knowledge merged with technological progress for the creation of worldwide innovations. The impact of technological change and globalisation has resulted in a massive formation of innovation through the use of new technologies and thus a division of industries into groups according to the degree of use of technology [34]. This creates layers of lower-tech industries, medium-high-tech and high-tech industries.

The role of low- and medium-technology companies and industries in modern economies is complex and frequently misunderstood [35]. In 1997, changes were made to the OECD classification of sectors and this began to accentuate only the high-tech industries. In practice, individual countries began to create specialised sectors for themselves that were able to use scientific knowledge and research, information and communication technologies, and often even biotechnology to generate innovation in its products [36]. Industrial manufacturing and other industries with a low ratio of technological components remain neglected; they are often moved to newly industrialised economies. This happens even when production in these industries is highly complex and capital intensive. A comparison of the index of industrial production in the major groups for the EU-27 between 2005 and 2012 is shown in Fig. 1.

Figure 1 provides evidence that low-tech industries have, from an economic perspective, high economic value, withstanding economic crises that are apparent in 2009. These findings are confirmed by a study [36]. When compared within the EU-27 in 2010 with other groups of technologically important sectors, the low-tech sector takes up 26 % (measured in added value at factor costs), medium-low, 27 %,

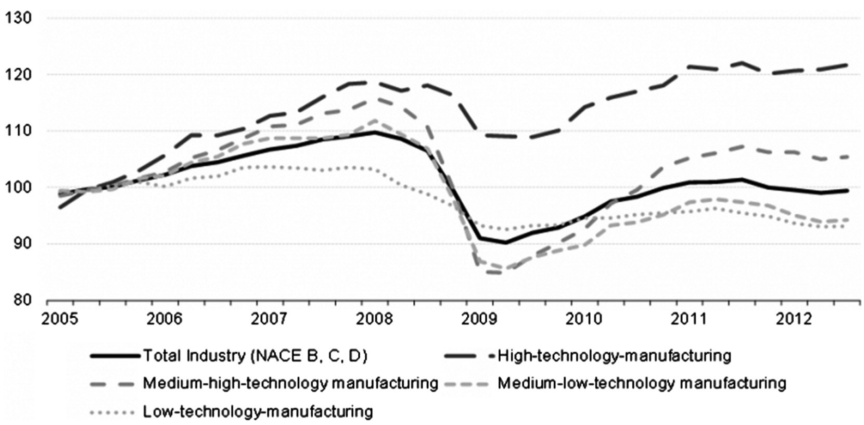


Fig. 1. Comparison of economic outputs of selected groups of technological sectors in 2012 in the EU-27. Source: http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/High-technology_versus_low-technology_manufacturing

medium-high 35 % and high-tech only 12 %. The actual contribution of individual industries to the total production of added value does not however show the dynamics of development. Available data from Eurostat shows that the index of average industrial production growth is continuously monitored from the individual sectors. This index, taking into account the period 2005–2010, indicates that in technologically intensive industrial sectors in the EU-27 there is a slight increase in the range of 1.0–3.0 %, while technologically weaker sectors on the other hand show a long-term slight decrease (to 1.0 %). In statistics comparing Member States, a very different development of these industrial sectors can be observed (see Table 1).

Table 1. Annual average growth in industrial sectors according to technological intensity (2005–2010) in selected countries.

Country	Technology level			
	High-tech	Medium-high	Medium-low	Low-tech
EU-27	3.3	1.0	−0.4	−0.7
Czech Republic	5.4	7.3	1.5	−1.4
Hungary	4.6	4.0	1.0	−1.1
Poland	14.5	8.4	6.9	3.0
Germany	6.6	2.3	1.8	0.1
France	1.4	−1.7	−1.8	−1.1
Portugal	−6.1	−3.1	−0.4	−1.5
Italy	−0.4	−2.2	−2.7	−1.8
Greece	−2.6	−5.7	−5.3	−2.9

Source: [37], Author’s own work

The table shows that even in the strongest economies in Europe (Germany and France), annual average growth in technologically strong sectors is not the same. Germany exceeds the European average in all four technology levels, while France has significantly fallen. In economically weak economies, often described with the acronym PIGS, development is very similar; all yearly values are falling. Among selected countries in Central Europe, it is evident that they belong to the group of countries where the emphasis on technology began to be emphasised within the last ten years, focusing on implementing measures of the so-called third wave mentioned above. Individual countries recorded significant annual growth in industrial sectors with high and medium-high technology levels.

Negative values for the "lower-tech" sector, however, clearly show that for all the monitored countries, these industries are no longer the centre of attention despite the fact that they constitute approximately 53 % of the added value of the annual production and are responsible for more than 30 % of the total exports of manufactured products [36]. Eurostat statistics, however, show that these industries are strategically important and bring significant economic value, and these companies are among the major employers. Attention on the part of the government serves also for their susceptibility to economic recession and crisis. In these periods, public intervention is necessary through appropriate economic policies so that businesses in technologically weak sectors can survive.

High-tech manufacturing industries also rank among the most significant group of companies. The largest number of high-tech industrial companies is located in Germany, Great Britain, Italy and the Czech Republic. In total, approximately 55 % of the high-sector is in the EU-27. It is worth noting that the UK has in recent years focused mainly on companies providing high-tech services based on knowledge (knowledge intensive services, KIS; [38]). This newly created KIS sector is recording higher growth in turnover and added value in nearly all EU-27 countries (except Finland, Slovakia and Hungary).

The importance of high-tech industries and KIS is also that this sector comprises small and medium companies, which can efficiently generate innovation and which use collaboration and networking to optimise the costs of creating new innovations, particularly in the area of science and research; they are more resistant to economic turbulence and rank among the major employers. This is confirmed by statistics: the number of people employed in manufacturing in the EU-27 in 2008–2011 decreased by 3.4 %, while in high-tech sectors if fell by only 2.5 % (in the given period, this can be explained by the economic crisis). This does not apply for all EU-27 countries, but only about 60 % of them. High-tech KIS also recorded in this period overall job growth of 1.3 %. In some EU countries, such as Slovakia and Luxembourg, a growth of employment in KIS was recorded in this period (from 7.5 to 16.5 %).

In the field of high-tech industries, it is necessary to mention the uneven distribution of companies in the country and thus significant regional disparities are often recorded [13]. Proximity to the capital or major regional centres has a particular influence here, as well as the location of major technical universities and research institutes. In the high-tech industry, the highest employment disparities are recorded between the most developed and the least developed region in Greece, Lithuania, Portugal and Romania, while the smallest differences are in Denmark, Ireland, Finland and Switzerland.

The importance of various high-tech-intensive companies can be seen on the basis of analyses of employment and selected macroeconomic indicators, of which Eurostat uses the ratio of employees, turnover ratio, product or added value, then gross investments in tangible goods. The importance is then analysed mainly through the influence of this group of companies on macroeconomic aggregates. The importance of various technologically important companies can be analysed according to their potential to contribute to growth through innovation, collaboration and networking.

3 Structural Equation Models of Knowledge Spillover Effects

Let X be a causal predictor of Y. Further, let M and W be moderators of the causal effect. Then the effect of X on Y may additively depend on both M and W and it may be expressed as follows [39]:

$$Y = i + (c_1 + c_4 \times M + c_5 \times W) \times X + c_2 \times M + c_3 W + e_Y, \tag{1}$$

where $(c_1 + c_4 \times M + c_5 \times W)$ denotes the conditional effect of X on Y. This effect may also depend multiplicatively on M and W. This can be described by the following equation:

$$Y = i + (c_1 + c_4 \times M + c_5 \times W + c_7 \times M \times W) \times X + c_2 \times M + c_3 W + c_6 \times M \times W + e_Y \tag{2}$$

where $(c_1 + c_4 \times M + c_5 \times W + c_7 \times M \times W)$ denotes the conditional effect of X on Y. In other words, the effect is a multiplicative function of M and W. In this case, a three-way interaction (moderated moderation) is present if c_7 is statistically different from zero.

Alternatively, the conditional effect of $X \times M$ may be moderated by W. This effect can be expressed in the following way:

$$Y = i + c_1 \times X + c_2 \times M + c_3 W + c_5 \times W \times X + c_6 \times M \times W + (c_4 + c_7 \times W) \times X \times M + e_Y, \tag{3}$$

where $(c_4 + c_7 \times W)$ is the conditional two-way interaction between X and M.

In a simple mediation model, X influences Y directly as well as indirectly through a single mediator variable M causally located between X and Y. The direct and indirect effects are then estimated from the following equations:

$$M = i_M + a_1 \times X + e_M \text{ and } Y = i_Y + c_1^{'} \times X + b_1 \times M + e_Y, \tag{4}$$

where c'_1 estimates the direct effect of X on Y, and $a_1 \times b_1$ estimate the indirect effect (see for example [40] for the methods of testing the indirect effect). The total effect c_1 can be then represented as a sum $c_1 = c'_1 + a_1 \times b_1$. More complicated models can be developed, including parallel and serial multiple mediator models, see [39] for details.

4 Data for Analysis

For the data collection we used a harmonised questionnaire of EU Member States from the Community Innovation Survey (CIS) carried out in the Czech Republic for the period 2008–2010. In total, data on 5,151 Czech companies with at least 10 employees was obtained (response rate greater than 60 %). We filtered only the companies in medium-high-tech and high-tech industries: 20–23 in NACE rev. 2. Also included were industries which the Eurostat classification system ranked as categories with a technology level.

The basic characteristics of the dataset are given in Table 2. The innovation activity of the industries was estimated by calculating the number of companies that introduced a new product or process to the market. Table 2 shows that there are significant differences between sectors.

Table 3 shows that innovative companies collaborate closely on the new product/process with other companies or institutions. The mode of cooperation on innovation seems to be similar for both types of industries, indicating that in-house cooperation prevails.

Table 4 shows the average degrees of importance of communication sources for innovation activities (it was measured on a scale from 0 – not used, to 3 – highly important). Internal communication was more important for innovative companies. Communication with customers, competitors, consultants, universities and the government was also more important to innovative companies. In contrast, communication with suppliers was less important. The situation with remaining sources was unambiguous.

Table 2. Average values of numerical determinants for high-tech industries.

Industry	Medium-high-tech		High-tech	
Innovative	NO	YES	NO	YES
TURN08	861,000	1,769,132	398,227	537,531
EMP08	7.0	520.0	449.4	307.3
RTOT10	12,889	78,555	10,616	22,484
N	533	307	134	51

Legend: TURN08 – total turnover in 2008, EMP08 – average number of employees in 2008, RTOT10 – total innovation expenditure in 2010.

Table 3. Percentage of companies collaborating on innovation activities.

Industry	Medium-high-tech		High-tech	
Innovative	NO	YES	NO	YES
COOP	28.9	55.4	27.3	52.9
IN-HOUSE		65.8		74.5
OTHFIRM		49.8		47.1
UNIV		22.5		27.5

Legend: COOP – cooperation on any of innovation activities, IN-HOUSE – innovation was developed by firm itself, OTHFIRM – together with other firms, UNIV – together with universities and other research institutions.

Table 4. Average degree of importance of communication sources for innovation activities.

Industry	Medium-high-tech		High-tech	
Innovative	NO	YES	NO	YES
SENTG	2.25	2.39	1.77	2.22
SSUP	1.92	1.83	1.82	1.65
SCLI	1.75	2.18	1.36	2.22
SCOM	1.40	1.65	1.23	1.80
SINS	0.72	0.88	0.64	0.80
SUNI	0.53	0.81	0.36	0.80
SGMT	0.36	0.50	0.36	0.41
SCON	1.20	1.56	1.55	1.47
SJOU	1.00	1.26	1.36	1.20
SPRO	0.75	0.94	1.00	0.80
SINT	1.61	1.76	1.82	1.73

Legend: SENTG – information within company, SSUP – information from suppliers, SCLI – information from customers, SCOM – information from competitors, SINS – information from consultants, SUNI – information from universities, SGMT – information from government, SCON – information from conferences, SJOU – information from scientific journals, SPRO – information from professional associations, SINT – information from Internet.

5 Results

The companies were asked to assign importance to communication sources: internal, market, institutional and other sources. For those with more than one source, a confirmatory factor analysis with maximum likelihood estimates was performed in the first step. The average weights (AW) for the four models (according to who developed the innovation – all options, by themselves, with other companies and with universities) are presented in Table 5.

Table 6 shows the results of the mediation analysis. Where a technologically oriented company creates innovation only through its own information and resources, it shows clear negative results regarding knowledge spillovers. We can therefore say that it results in a paradoxical form of negative knowledge spillovers (in terms of an

Table 5. Confirmatory factor analysis and reliability analysis on communication sources.

Internal	AW	Market	AW	Institutional	AW	Other	AW
within firm	1	Suppliers	.407	Universities	.406	Conferences	.624
		Customers	.742	Research institutes	.525	Scientific journals	.842
		Competitors	.805			Professional associations	.616
		Consultants	.303			Internet	.571

Source: Author's own work

Table 6. Mediation models for internal/external communication.

Effect	Innovation developed			
	All options	By itself	With companies	With universities
Total effect – internal	.261*	−.134	.506***	.294*
Direct effect – internal	.176	−.105	.463***	.195
Indirect effect – internal	.102***	−.029	.069*	.143**
Total effect – market	.457***	−.062	.123	.124
Direct effect – market	.339***	−.010	.002	−.095
Indirect effect – market	.135***	−.053*	.124***	.252***
Total effect - instit.	.372**	.361**	.051	1.135***
Direct effect - instit.	.120	.561***	−.242*	.879***
Indirect effect - instit.	.247***	−.191***	.287***	.372***
Total effect – other	.282**	.461***	−.057	.660***
Direct effect – other	.135	.579***	−.242*	.469***
Indirect effect – other	.158***	−.107***	.172***	.284***

Legend: *significant at $P < 0.10$,** significant at $P < 0.05$ *** significant at $P < 0.01$.
Source: Author's own work

analogy to the so-called negative externalities known from the theory of public economics). This result was observed most in the analyses of indirect effects (institutional source of information −.0191*** and other source of information −.107***). In contrast, in all other cases the existence of positive knowledge spillovers was clearly demonstrated. The smallest are where businesses draw information from internal sources. On the contrary, the largest indirect effect was observed in obtaining institutional information (from universities and research institutions). The largest direct effect on innovation activities then is that of market information.

Abramovsky [41] reported that financial constraints may be positively related to undertaking collaborative innovation. Therefore, we also tested the moderating effects of (1) expenditure on R&D (two categories created using median as a separator), (2) government support, and (3) EU support, see Table 7. An influence on R&D costs were seen, which were influenced by institutional and other information sources, and an influence on both innovation developed by itself and in cooperation with universities was also discovered. In the case of influence on innovation developed with other companies and collaboration activity, this is an interaction with internal information sources. A major finding is evidence of the influence of utilising government support and other subsidies from the EU. Utilisation of government support and support from the EU has no effect on knowledge spillovers in high-tech industries.

Table 7 shows significant findings. Analysis showed that if companies carry out private investment in the development of innovations, a certain degree of spillover effects will occur. In the event that these investments complement cooperation with R&D organisations or universities, the value of the resulting spillover effects increases. Another situation arises after analysing public government spending. Accepting government support for businesses encourages greater cooperation, but the resulting volume of spillover effects created is lower. This may be due to a time lag or less

Table 7. Moderation models for internal/external communication.

Interaction	Cooperation	Innovation developed			
		All options	By itself	With companies	With universities
RTOT10 × internal	.508**	−.088	−.359	.485**	.172
RTOT10 × market	.058	−.018	.030	−.014	.377
RTOT10 × instit.	−.038	−.222	.571*	−.098	.439
RTOT10 × other	−.113	−.346	.547*	−.114	.723**
GOVER × internal	−.091	.537	.328	−.352	−.801**
GOVER × market	.096	−.201	.128	−.288	−.265
GOVER × instit.	.164	.411	−.023	−.391	−.013
GOVER × other	.384	.477	.127	−.241	−.089
EU × internal	.773	.085	.388	−.132	−.452
EU × market	−.157	−.607	−.184	−.166	−.470
EU × instit.	.078	−.099	−.492	−.105	−.304
EU × other	−.049	−.063	−.454	−.256	−.411
RTOT10 × coop.	–	−.220	−.053	.869*	.546
GOVER × coop.	–	.418	.366	−.648	.235
EU × coop.	–	.769	−.201	−.073	.292

Source: Author's own work

pressure on the efficient utilisation of such allocated funds, for example, by the influence of the degree of bureaucracy. The situation is similar for EU subsidies. It has been shown that European funds also support collaboration, but the extent of the resulting spillover effects is still lower than that of domestic government funding.

The analysis has also provided evidence that private funds utilised for collaboration between companies lead to the creation of innovation (.869*). While the use of public funds by the companies themselves leads to the formation of innovation to a lesser extent (.366) than if it were to seek cooperation with universities and R&D organisations (.235).

6 Conclusion

The aim of this paper was to carry out an evaluation of the importance of industrial sectors which have a higher technological level and at the same time to analyse the spillover effects that arise in the creation of innovation - on the example of Czech companies. With the aid of moderated and mediation analysis, evidence was shown that the influence of cooperation would create spillover effects, but did not confirm the assumptions that collaboration between companies and the scientific research sector generates higher spillover effects. We have shown that the influence of collaboration creates spillovers that facilitate innovative activity. Evidence was also shown that the influence of various factors can give rise to negative spillover effects. This can be possibly explained by a low absorptive capacity of Czech high-tech industries [42].

Furthermore, it was found that the most direct effect on innovative activity is the acquisition of information from current markets.

The actual quantification of the effects which influence collaboration is important for further analysing the effectiveness and the possible formation of public policies and subsidy schemes. It has been clearly demonstrated that drawing government aid and aid from EU funds helps to produce cooperative links between different organisations. A significant influence on knowledge spillover effects has however not been confirmed in high-tech industries.

Based on these results, it is possible to advise public officials to use public funds to promote cooperation between businesses and various economic organisations in the direction of creating innovations, as well as forming a platform for the transfer of information, technology or know-how. Rules for such support must however be focused on achieving effectiveness and must not create bureaucratic obstacles that result in reduced effectiveness of the use of public funds. Specifically, we recommend collaborative scientific and research projects between business and academia such as PhD theses.

Resolving this complex research project and the interim results invite further questions. We will try to resolve these in further research where more complicated models can be developed, including parallel and serial multiple mediator-moderator models. We also encourage future research in examining various methods for standardized indirect effects' analysis.

Acknowledgments. This article was created as part of the resolution of the research task No. 14-02836S entitled "Modelling of knowledge spillover effects in the context of regional and local development", financially supported by the Grant Agency of the Czech Republic.

References

1. Daly, H.E.: Ecological Economics and the Ecology of Economics: Essays in Criticism. Edward Elgar, Cheltenham (1999)
2. Gerybadze, A., Reger, G.: Globalization of R&D: recent changes in the management of innovation in transnational corporations. Res. Policy **28**(2), 251–274 (1999)
3. Lambin, E.F., Meyfroidt, P.: Global land use change, economic globalization, and the looming land scarcity. Proc. Natl. Acad. Sci. **108**(9), 3465–3472 (2011)
4. Aggarwal, R.: Developing a global mindset: integrating demographics, sustainability, technology, and globalization. J. Teach. Int. Bus. **22**(1), 51–69 (2011)
5. Fujita, M., Thisse, J.F.: Economics of Agglomeration: Cities, Industrial Location, and Globalization. Cambridge University Press, Cambridge (2013)
6. Meinhard, S., Potrafke, N.: The globalization – welfare state nexus reconsidered. Rev. Int. Econ. **20**(2), 271–287 (2012)
7. Stejskal, J., Hajek, P.: Competitive advantage analysis: a novel method for industrial clusters identification. J. Bus. Econ. Manag. **13**(2), 344–365 (2012)
8. Leigh, N.G., Blakely, E.J.: Planning Local Economic Development: Theory and Practice. SAGE Publications Incorporated, London (2013)
9. Deininger, K.: Challenges posed by the new wave of farmland investment. J. Peasant Stud. **38**(2), 217–247 (2011)

10. Arrow, K.J., Lind, R.C.: Uncertainty and the evaluation of public investment decisions. J. Nat. Resour. Policy Res. **6**(1), 29–44 (2014)
11. Eisner, M.A.: The American Political Economy: Institutional Evolution of Market and State. Routledge, London (2013)
12. Lee, S.M., Hwang, T., Choi, D.: Open innovation in the public sector of leading countries. Manag. Decis. **50**(1), 147–162 (2012)
13. Hansen, T., Winther, L.: Innovation, regional development and relations between high-and low-tech industries. Eur. Urban Reg. Stud. **18**(3), 321–339 (2011)
14. Matatkova, K., Stejskal, J.: Descriptive analysis of the regional innovation system-novel method for public administration authorities. Transylvanian Rev. Adm. Sci. **39**, 91–107 (2013)
15. Hajek, P., Henriques, R., Hajkova, V.: Visualising components of regional innovation systems using self-organizing maps - evidence from European regions. Technol. Forecast. Soc. Change **84**, 197–214 (2014)
16. Braguinsky, S., Klepper, S., Ohyama, A.: High-tech entrepreneurship. J. Law Econ. **55**(4), 869–900 (2012)
17. Chesbrough, H.: Business model innovation: opportunities and barriers. Long Range Plan. **43**(2), 354–363 (2010)
18. Camagni, R., Capello, R.: Regional Innovation patterns and the EU regional policy reform: toward smart innovation policies. Growth Change **44**(2), 355–389 (2013)
19. Trigo, A., Vence, X.: Scope and patterns of innovation cooperation in spanish service companies. Res. Policy **41**(3), 602–613 (2012)
20. Zeng, S.X., Xie, X.M., Tam, C.M.: Relationship between cooperation networks and innovation performance of SMEs. Technovation **30**(3), 181–194 (2010)
21. Cao, M., Zhang, Q.: Supply chain collaboration: impact on collaborative advantage and firm performance. J. Oper. Manag. **29**(3), 163–180 (2011)
22. Cellini, R., Lambertini, L.: Dynamic R&D with spillovers: competition vs. cooperation. J. Econ. Dyn. Control **33**(3), 568–582 (2009)
23. Villasalero, M.: Signaling, spillover and learning effects of knowledge flows on division performance within related diversified firms. J. Knowl. Manag. **17**(6), 928–942 (2013)
24. Frenz, M., Ietto-Gillies, G.: The impact on innovation performance of different sources of -knowledge: evidence from the UK community innovation survey. Res. Policy **38**(7), 1125–1135 (2009)
25. Belderbos, R., Carree, M., Lokshin, B.: Cooperative R&D and firm performance. Res. Policy **33**(10), 1477–1492 (2004)
26. Cassiman, B., Veugelers, R.: R&D cooperation and spillovers: some empirical evidence from Belgium. Am. Econ. Rev. **92**(2), 1169–1184 (2002)
27. López, A.: Determinants of R&D cooperation: evidence from Spanish manufacturing firms. Int. J. Ind. Organ. **26**(1), 113–136 (2008)
28. Montoro-Sánchez, A., Ortiz-de-Urbina-Criado, M., Mora-Valentín, E.M.: Effects of knowledge spillovers on innovation and collaboration in science and technology parks. J. Knowl. Manag. **15**(6), 948–970 (2011)
29. Andreeva, T., Kianto, A.: Knowledge processes, knowledge-intensity and innovation: a moderated mediation analysis. J. Knowl. Manag. **15**(6), 1016–1034 (2011)
30. Simonen, J., McCann, P.: Innovation, R&D cooperation and labor recruitment: evidence from Finland. Small Bus. Econ. **31**(2), 181–194 (2008)
31. Un, C.A., Romero-Martínez, A.M., Montoro-Sanchez, A.: Determinants of R&D collaboration of service firms. Serv.Bus. **3**(4), 373–394 (2009)
32. Mendonça, S.: Brave old world: accounting for 'high-tech' knowledge in 'low-tech' industries. Res. Policy **38**(3), 470–482 (2009)

33. Murmann, J.P.: The Coevolution of Industries and Comparative Advantage: Theory and Evidence. Cambridge University Press, Cambridge (2003)
34. Evangelista, R., Vezzani, A.: The economic impact of technological and organizational innovations. A Firm-Level Anal. Res. Policy **39**(10), 1253–1263 (2010)
35. Robertson, P., Smith, K.H., Von Tunzelmann, N.: Innovation in low-and medium-technology industries. Res. Policy **38**(3), 441–446 (2009)
36. Hirsch-Kreinsen, H., Jacobson, D., Robertson, P.L.: 'Low-tech' industries: innovativeness and development perspectives - a summary of a European research project. Prometheus **24**, 3–21 (2006)
37. Jaegers, T., Lipp-Lingua, C., Amil, D.: High-Technology and Medium-High Technology Industries Main Drivers of EU-27's Industrial Growth. Statistics in Focus 1 (2013)
38. Von Nordenflycht, A.: What is a professional service firm? Toward a theory and taxonomy of knowledge-intensive firms. Acad. Manag. Rev. **35**(1), 155–174 (2010)
39. Hayes, A.F.: A versatile computational tool for observed variable mediation, moderation, and conditional process modelling [White paper] (2012). http://www.afhayes.com/public/process2012.pdf
40. Williams, J., MacKinnon, D.P.: Resampling and distribution of the product methods for testing indirect effects in complex models. Struct. Equ. Model. **15**(1), 23–51 (2008)
41. Abramovsky, L., Kremp, E., López, A., Schmidt, T., Simpson, H.: Understanding co-operative innovative activity: evidence from four European countries. Econ. Innov. New Technol. **18**(3), 243–265 (2009)
42. Spithoven, A., Clarysse, B., Knockaert, M.: Building absorptive capacity to organise inbound open innovation in traditional industries. Technovation **31**(1), 10–21 (2011)

Indigenous Knowledge and Socioeconomic Development: Indigenous Tourism in Kenya

Tom Kwanya(✉)

Department of Information and Knowledge Management,
The Technical University of Kenya, P.O. Box 52428, Nairobi 00200, Kenya
tom.kwanya@gmail.com

Abstract. Indigenous knowledge plays a pivotal socioeconomic development role in indigenous communities. In Kenya, one of the economic sectors where indigenous knowledge can be applied is tourism which is among the country's major income earners besides tea and horticulture. This study investigated the potential and the actual use of indigenous knowledge in leveraging the other efforts being made to develop and cushion tourism in Kenya.

The study was designed as a survey to capture the current status of the application of indigenous knowledge in the tourism sector in Kenya. Primary data was collected through key informant interviews with tourism industry stakeholders in the country. 56 participants in the study were selected through a mix of stratified, purposive and snowballing sampling techniques. The interviews were conducted using semi-structured questionnaires administered by the researcher. Secondary data on indigenous knowledge in Kenya, indigenous tourism as well as tourism sector statistics was collected through documentary analysis. The data was analysed using content analysis.

The findings of the study indicate that although indigenous tourism holds a great socioeconomic potential in Kenya, it has not been harnessed fully. Its potential is still being held back by myriad challenges such as lack of relevant business development skills; lack of adequate capital to develop and promote indigenous tourism products, services and facilities; remoteness of indigenous tourism sites; insecurity; poor infrastructure; modernisation; environmental degradation and consequences of climate change; stiff competition; and intra or inter-ethnic resource-based conflicts. The findings can be used by the Government of Kenya to mainstream indigenous knowledge into tourism by developing the requisite policies, structures and implementation frameworks. The findings may also be used by the tourism sector stakeholders in Kenya to identify, enhance, package and promote indigenous tourism products, services and facilities effectively.

Keywords: Indigenous knowledge · Knowledge management · Kenya · Tourism · Socioeconomic development

1 Introduction

Indigenous knowledge (IK) is defined as the unique, traditional, local knowledge existing within and developed around the specific conditions of people indigenous to a particular geographic area [8]. Boven and Morohashi [2] also define IK as the local

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 342–352, 2015.
DOI: 10.1007/978-3-319-21009-4_26

knowledge which is unique to a given culture and society. IK can also be perceived as the knowledge, innovations and practices of indigenous communities in matters related to agriculture, environmental management, medicine and health, art and language. Therefore, IK is the anchor for survival and stability in indigenous communities and acts as the information base for local-level decision making in agriculture, health care, food preparation, education, natural-resource management, and a host of other activities in rural settings [16]. IK is about the ways of knowing, seeing, and thinking that are passed down orally from generation to generation, and which reflect thousands of years of experimentation and innovation in all aspects of life in a particular context. Ocholla and Onyancha [12] also describe IK as a dynamic archive of the sum total of knowledge, skills and attitudes belonging to a community over generations and expressed in the form of action, object and sign languages for sharing. However, they point out that for reasons largely associated with ignorance and arrogance, indigenous knowledge has unfortunately been neglected, vindicated, stigmatised, legalised and suppressed among the majority of the world's communities. IK is sometimes also referred to as Indigenous Knowledge Systems and Practices (IKSP) or Indigenous Technological Knowledge (ITK).

Indigenous knowledge has value not only for the culture in which it develops, but also for the other stakeholders. The World Bank [17] explains that IK provides the basis for problem-solving strategies for local communities, especially the poor, and represents an important component of global knowledge on development issues. Nonetheless, the World Bank [17] also points out that IK is an underutilised resource in the development process. The Bretton Woods institution suggests that learning from IK by investigating first what local communities know and have can improve the understanding of local conditions; provide a productive context for activities designed to help the communities; and increase responsiveness to issues. It further proposes that adapting international practices to the local setting can help improve the impact and sustainability of development assistance; sharing IK within and across communities can help enhance cross-cultural understanding and promote the cultural dimension of development; and most importantly, investing in the exchange of IK and its integration into the assistance programs of development partners can help to reduce poverty.

Although IK systems contrast remarkably with scientific conventional knowledge systems, they are dynamic and are also influenced by internal creativity and experimentation as well as by contact with external systems [3]. The value of IK increases with its applicability to address emerging challenges facing communities. For instance, its applicability as an enabler of societal socioeconomic development is one of the yardsticks with which its value to the communities, especially in modern times, may be measured.

2 Theoretical Framework

The author used the knowledge worldview model outlined by Fleer [4] as the theoretical framework for this study. Hart [9] defines knowledge worldviews as cognitive, perceptual, and affective maps that people continuously use to make sense of their social landscape and to find their ways to whatever goals they seek. He further explains that knowledge worldviews are developed throughout a person's lifetime through socialisation and social interaction. He points out that they are encompassing and pervasive in adherence and influence yet they are usually unconsciously and uncritically taken for granted as the way things are.

The model proposed by Fleer [4] classifies perceptions of knowledge into two broad categories, that is, indigenous and scientific knowledge. She explains that these sets of knowledge define one's worldviews. She points out that while indigenous knowledge is founded on traditional worldview and produced for specific purposes to maintain particular societies, scientific knowledge is founded on the "civilised" worldview and produced for the sake of it. She argues that while scientific knowledge seeks power over nature and people, indigenous knowledge seeks to coexist with the same. She describes scientific knowledge as being materialistic, reductionist, rational, decontextualised, individual and competitive. On the other hand, she extols indigenous knowledge as being spiritual, holistic, intuitive, contextualised, communal and cooperative.

While applying this model, the researcher is, however, aware that although there appears to be a clear dichotomy between indigenous and scientific knowledge, this division is not realistic since these sets of knowledge interact with and influence each other. For instance, indigenous knowledge can be investigated, validated and documented in the same way as scientific knowledge. Thus, indigenous knowledge can produce scientific knowledge and vice versa. This illustrates the closeness of the two sets of knowledge.

The knowledge worldviews model was applied to help the researcher to understand what really constitutes indigenous knowledge as well as its influence on its adherents' perception of the world around them. The researcher also used the model to unravel the actual and potential uses of indigenous knowledge to improve and sustain livelihoods through socioeconomic activities such as tourism in rural communities.

3 Study Rationale

The United Nations World Tourism Organization (UNWTO) in 2006 estimated that different countries worldwide earned a total of USD 2 billion daily from tourism related services and products. This revenue accounted for 6 percent of the world's total export earnings [14] making tourism the fourth biggest export earner after fuels, chemicals and automobile products. Valle and Yobesia [15] explain that the significant contribution of the tourism sector to most economies is generated in terms of increased foreign exchange receipts, balance of payments, government revenues, employment, and increased economic activity in general.

Kenya is one of the countries whose economies rely immensely on tourism. Indeed, tourism is one of the major income earners in Kenya besides tea and horticulture. Valle and Yobesia [15] explain that the position of tourism as an important foreign exchange earner in Kenya has grown over the years with the decline of prices of agricultural commodities in the world markets. The latest available statistics from the Government of Kenya indicate that the country had attracted 1,095,945 tourists and earned 73.68 billion Kenya Shillings (about USD 1 billion according to the exchange rate) in 2010.

Nonetheless, this critical source of national revenue is hampered by local insecurity as well as piracy in the Indian Ocean affecting the number of cruise ships docking in the country's ports; rising incidents of terrorism-related attacks leading to unfavourable travel advisories; socioeconomic challenges such as Euro Zone in its traditional tourist baskets; poaching of wildlife; as well as inadequate infrastructure and accommodation facilities in some areas. The impact of these challenges is serious. Valle and Yobesia [15] report that Kenya's share of the world's international tourist arrivals rose from 0.17 % in 1985 to 0.19 % in 1990 but dropped to 0.12 % in 2001. A similar trend was witnessed with the country's share of Africa's tourist visits which dropped from 4.7 % in 1990 to 3.2 % in 2000. They further explain that Kenya was the sixth most important tourist destination in Africa in 2000 but had been overtaken by Algeria and Nigeria in 2001.

The statistics presented above demonstrate that the tourism sector in Kenya is under siege. Whereas the Government of Kenya has put in place myriad strategies to stabilise the tourism sector in the country its fortunes keep dwindling by the day. For instance, the Kenya Tourism Federation[1], citing statistics from the Kenya National Bureau of Statistics, reported that 2014 was one of the most challenging years for the tourism sector in Kenya evidenced by the 14.6 % decline in the sector's performance in the 3rd quarter of the year, up from a 3.9 % decline during a similar period in 2013. Evidently, there is need to develop new strategies to enrich and promote the tourism sector in Kenya. Indigenous knowledge is one of the ways by which new tourist attraction streams may be realised. In spite of its great potential in promoting tourism, indigenous knowledge is still not mainstreamed in the tourism sector in Kenya. There is need, therefore, to demonstrate the real potential of indigenous knowledge to support socioeconomic development in Kenya through indigenous tourism.

4 Methodology

This study investigated the forms of indigenous knowledge which have a great potential to contribute to tourism in Kenya; the current status of the use of indigenous knowledge to attract tourists and boost revenues from tourism in Kenya; and the challenges hampering the effective mainstreaming of indigenous knowledge in the tourist corridors in Kenya. The study also explored and recommended strategies that may be used to mainstream indigenous knowledge in the tourism sector in Kenya.

The study was designed as a survey to capture the current status of the application of indigenous knowledge in the tourism sector in Kenya. Primary data was collected through key informant interviews with tourism industry players in Kenya. 56 participants in the study were selected through a mix of stratified, purposive and snowballing sampling techniques. First, the researcher identified five major clusters of stakeholders in the tourism sector in Kenya. These were owners of tourist facilities like hotels and lodges; direct employees such as tour guides; travel agents and ticketing service providers; producers of diverse handicrafts; and the relevant government agencies

[1] Report obtained from http://www.ktf.co.ke/news.asp?id=115 on 17 February 2015.

and officials. The researcher thereafter selected ten respondents from each cluster through information-oriented purposive sampling. Additional six respondents were identified through snowballing, that is, as recommended by the other respondents. The interviews were conducted using semi-structured questionnaires administered by the researcher. Secondary data on indigenous knowledge in Kenya, indigenous tourism as well as tourism sector statistics was collected through documentary analysis. The data was analysed using content analysis.

5 Findings and Discussions

The findings of the study are presented and discussed in the sections hereunder.

5.1 Potential Role of IK in Tourism in Kenya

Kenya's tourism can benefit immensely from IK through indigenous tourism. Also known as cultural tourism, indigenous tourism can be defined as a tourism activity in which the indigenous people are directly involved either through control and/or by having their culture serve as the essence of the attraction [10, 18]. Indigenous tourism can be achieved, for instance, by restaurants which serve indigenous foods; hotels and accommodation facilities owned and managed by indigenous people or constructed using indigenous architecture; indigenous games and cultural events; lifestyles of indigenous communities; indigenous art exhibitions, music, dances and stories around campfires; cultural ceremonies, festivals and special events; visits to cultural sites and shrines; cultural, environmental and spiritual beliefs and practices of indigenous people; and museums holding indigenous artefacts. Smith [13] states that habitat, heritage, history and handicrafts are the four "Hs" which underpin indigenous tourism. Besides creating jobs for local people, indigenous tourism also gives indigenous communities an opportunity to tell their story to the world. Indigenous tourism gives the tourists an authentic experience with the indigenous communities as well as their culture and environment which is not possible to get through any other way. Indigenous tourism is one of the fastest growing tourism niches in the world today.

Kenya has several attractions which can be exploited for indigenous tourism. Some of these attractions include the six UNESCO World Heritage sites in the country. The six are Mount Kenya National Park and natural forest, Fort Jesus, Old Town Lamu, Kenya lake system in the Great Rift Valley, the sacred Mijikenda Kaya Forest and last but not least the Turkana National Park. It is a great honour for Kenya to play host to these six sites which are famous internationally for their ecological and cultural treasures [11]. Kenya is also home to several prehistoric sites which can be a great indigenous tourism attraction. One of these is Koobi Fora where one of the earliest evidences of human habitation on earth was found. Other prehistoric and archaeological sites in Kenya include Olorgesailie, Enkapune Ya Muto, Hyrax Hill, Jumba la Mtwana, Ruins of Gedi, Manda Island, Namoratunga and Pate Island.

Kenya also has a number of sacred forests which can be major indigenous tourism attractions. Adam [1] provides an overview of some of the sacred natural sites in

Kenya. These include Karima Forest which is considered as an *ihoero* (sacred natural site in the local Kikuyu culture). It is located in Othaya division of Nyeri County in central Kenya. It is a tapering dome-shaped volcanic hill with its highest point being at an altitude of 6000ft above sea level. It is located between the sacred Kirinyaga mountain (Mt. Kenya) and the Nyandarua (Aberdare) ranges about 150 km north-east of the capital, Nairobi. There are two shrines, Kamwangi and Gakina in Karima Forest, comprising 85 acres, which are gazetted under the National Museums of Kenya. The other sacred site is the Mijikenda Kaya forests which are the most well-known of Kenya's cultural heritage sites. The area consists of several forest sites spread over 200 km in the contiguous Kenyan coastal counties of Kwale, Mombasa, Kilifi and Malindi. The Mijikenda people respect the Kaya forests as the abodes of their ancestors and are revered as sacred natural sites. The sacred natural sites owe their continued existence largely to the cultural knowledge and practices of the nine coastal Mijikenda ethnic groups - the Giriama, Digo, Duruma, Rabai, Kauma, Ribe, Jibana, Kambe and Chonyi. Another sacred forest is Giitune which is considered as an *irii* (sacred natural site in local culture) on the eastern side of Mt. Kenya and is one of the numerous sacred natural sites surrounding this UNESCO World Heritage site. Giitune lies in a high rainfall area with fertile and well-drained volcanic soils. It is a community forest under the governance of the community and is "heritage" recognised and protected under the National Museums of Kenya. Still there are Mathembo sacred natural sites in Ukambani in eastern Kenya where the Kamba community offers sacrifices during droughts and epidemics or to give thanks for a good harvest. The trees and bushes growing in these places are highly protected and cutting them is prohibited. Kivaa sacred natural site in Masinga is an example of an *ithembo* (sacred natural site in local Kamba language) which has been rehabilitated by the local community through a revival of their cultural practices and governance systems related to ecosystems.

Forty-two ethnic groups living in Kenya with their diverse cultural practices offer tourists a unique experience. Right from traditional dances, pastoral life and bull fighting to polygamous practices, festivals and celebrations, Kenyan cultural life is rich, offers academic and study tourists a fascinating destination. The myths surrounding "kaya" in the coast, "Kit Mikayi" and "Simbi Nyaima" in the west among others are potentially great attractions. Several museums preserving national cultural heritage are spread across the country: Nairobi, Kisumu, Kitale, Kapenguria, Lamu and Meru. Others are Fort Jesus in Mombasa, Gedi Ruins in Malindi, Thimlich Ohinga in South Nyanza, and Jumba la Mtwana ruins in Kilifi. Every corner of Kenya has unique cultural features of tourist interest [6].

5.2 Current Status of the Application of IK in Tourism in Kenya

The majority (86 %) of the respondents acknowledged that Kenya has not exploited its indigenous tourism potential fully. They explained that the bulk of Kenya's tourists are currently attracted by the wilderness and wildlife as well as the sandy beaches; not indigenous tourism factors. The respondents pointed out that the government does not seem keen to harness Kenya's great indigenous tourism potential. They explained that this view is supported by the fact that government does not seem to prioritise indigenous tourism. Indeed, the Government of Kenya's 2013–2018 tourism strategy lists natural resources and cultural resources as the 13th and 14th pillars of attraction respectively [7]. These pillars are actually listed last. This perhaps indicates the low opinion with which the government considers the potential of indigenous tourism in the country.

The other respondents (14 %) were of the view that the government is doing its best to integrate IK into its tourism strategy. They argued that it is not possible to separate indigenous tourists from the other forms of tourism because the experience is intricately intertwined. They explained that while many tourists visit Kenya to experience its famous wildlife or sandy beaches, for many more the local culture makes their stay special. They further referred to exit polls among departing guests at airports which show their appreciation for the warm and welcoming spirit of the Kenyan people. These respondents seemed to echo the words of Gachie [5] that a trip to Kenya is about more than just wildlife or scenery - the real face of Kenya is found among the combined faces of its many cultures. It is the people who bring the destination to life – each landscape has a different cultural significance to a different community. These respondents also pointed out that wildlife has long been an essential part of traditional Kenyan culture ensuring the conservation of wild animals and ecosystems. They also referred to the cultural centres and activities including the Bomas of Kenya, Maulidi festival in Lamu and a couple of other indigenous attractions which have been developed and promoted by the government as an indication that some efforts have been made to harness the potential of IK to enrich Kenya's tourism appeal and competitiveness.

Considering the two schools of thought expressed by the respondents above, it is evident that some effort is being made to embrace IK as a tourist attraction. Nonetheless, there seems to be no structured, documented strategy to identify, develop, market or manage indigenous tourism attractions and destinations in Kenya.

5.3 Challenges of Applying IK in Tourism in Kenya

The respondents identified several challenges which hamper the effective application of IK in tourism in Kenya. These include lack of adequate capital to develop products, market the attractions, build facilities and hire competent staff. The other challenge is lack of relevant business skills amongst local communities to develop appropriate business models around IK and manage indigenous tourism ventures. Remoteness of indigenous tourist attraction sites is also another challenge which pushes up costs associated with visiting the attraction sites. The remoteness of some of the indigenous tourism territories makes them vulnerable to insecurity and lack of basic services essential for hospitality. Given that some of the attractions are based on natural resources, the effects of climate change and the resulting environmental degradation negatively affect the viability of indigenous tourism in those areas.

Indigenous tourism is also hampered by a poor understanding of it amongst the communities leading to an equally poor attitude towards it. The potential of indigenous tourism is also held back by civilisation which has eroded some of the cultural attributes of communities and thus making them ordinary. In some cases indigenous tourism is hampered by fake practitioners. One of the common cases involves fake Maasai *morans* (warriors) especially at the coast. Once the customers realise that they have been conned, they become repugnant to such practices, authentic cases.

The potential of indigenous tourism in Kenya is also affected by inter or intra-communal conflicts over natural resources and other issues. These conflicts not only lead to insecurity in the affected areas but also leave the potential customers confused. Sometimes the conflicts lead to the destruction of the attraction sites or resources. Recent examples include the 2007–2008 post-election ethnic violence; invasion of the Kaya forests by Mombasa Republican Council insurgents; terrorism attacks in Malindi and Lamu; as well as political conflicts resulting from the ownership wrangles over the Maasai Mara by Narok County government and the local members of parliament.

Another challenge facing indigenous tourism in Kenya is competition from other destinations in the region. South Africa, Rwanda and Tanzania are some of the indigenous tourism destinations which present a stiff competition to Kenya. This competition is exacerbated by poor packaging and marketing of indigenous tourism attractions in Kenya.

5.4 Strategies to Best Apply IK in Tourism in Kenya

The aforementioned challenges may be addressed through the following as suggested by the respondents or gleaned from secondary data:

(1) The stakeholders, especially the government players, should develop a clear strategy on indigenous tourism. The strategy should identify the high potential niches and outline the means of harnessing them. Such a strategy should be both short-time and long-term.
(2) The Government of Kenya should create structures and institutional frameworks facilitative of indigenous tourism. One of such structural frameworks could be a directorate dealing with indigenous tourism. These structures should percolate down to the devolved levels of government where most attractions are located.
(3) The stakeholders should develop unique indigenous tourism products which can compete globally. They should not just imitate what their competitors are doing but must make efforts to develop and repackage products which are unique to Kenya.
(4) The stakeholders should also market indigenous tourism attractions and destinations aggressively. This can be done through expos, demos, publications, documentaries, goodwill ambassadors and other forms of promotion regionally, nationally and globally.
(5) The stakeholders should mobilise adequate resources to capitalise indigenous tourism activities. The capital can be used to package products, promote attractions and destinations, hire competent staff, preserve and conserve cultural practices, and to build requisite facilities. The capital can be raised through communal mobilisation, financial institutions, donors and other philanthropic entities, and government funding.

(6) The Government of Kenya should develop the infrastructure essential for indigenous tourism. The basic infrastructure includes telecommunication, transportation, accommodation and general services. Whereas, it is essentially the responsibility of the government to develop the infrastructure, the affected communities should also mobilise resources to complement the government efforts. The communities should ensure that the available infrastructure is used well to enhance its longevity and utility. The communities should also protect the infrastructure against unwarranted damage and vandalism.
(7) The communities should work together with the government and other stakeholders to enhance security not only to protect indigenous tourism attractions but to keep the communities peaceful and governable. The current insecurity situation in Kenya is so grave that most of the tourist sources have issued advisories to their citizens against travelling to the country. In fact, some tourist facilities have closed resulting in job and revenue losses for thousands of people.
(8) Indigenous tourism relies greatly on cultures and natural habitats. The concerned communities should work to conserve and preserve these to continue being attractive to tourists. The conservation of natural resources and preservation of relevant cultures require concerted efforts of diverse stakeholders. The government and community leaders should rally the efforts of all the people to make this possible.
(9) Another strategy of encouraging indigenous tourism is building networks and alliances necessary to enhance the environment in which the trade occurs. These networks could be local, regional, national or international. They may be used for advocacy on issues pertinent to indigenous tourism, resource mobilisation, policy formulation and implementation as well as marketing of attractions and destinations. The bigger the networks and alliances the better. Nonetheless, the participants should ensure they are strong enough to augment their impact.
(10) The Government of Kenya should make efforts to mainstream indigenous tourism. The fact that it is included in the national 2013–2018 strategy is encouraging. However, the fact that it is listed at the tail end of the strategy is a cause of concern as it may be an indicator that the government does not take indigenous tourism as seriously as it takes the other forms of tourism. It should not be perceived and presented as an option only pursued when the other options are not tenable.

6 Conclusion

Indigenous knowledge is an important facilitating factor to socioeconomic development of indigenous communities. Indeed, indigenous knowledge grew out of the need of communities to survive and thrive in particular environments. Its role in cushioning communities against economic shocks and ensuring sustainable development and stability cannot be overstated. In a world which is steadily modernising, tourism is one of the areas which indigenous knowledge can make a significant socioeconomic contribution. Used properly, it cannot just save lives and households but also provide a critical development pedestal upon which other economic initiatives can be launched.

Although the potential of indigenous tourism in Kenya seems to have been recognised as evidenced by the establishment of cultural centres like the Bomas of Kenya, it has been treated as a peripheral factor in tourism development in Kenya. Its latent value for tourist attraction include its sustainable nature, competitive advantage due to its uniqueness to the country, ability to empower the marginalised communities, capacity to open up remote areas of the country as well as support for climate conservation and preservation. The national Government of Kenya as well the county governments should make deliberate efforts to harness this potential. Significantly, they should develop a strategy to map out indigenous tourist attractions as well as improve and market them alongside the other attractions. Importantly, indigenous tourism should not be treated merely as a pastime which is only pursued when the other options are not possible.

7 Implications of the Findings of the Study

These findings can be used by the Government of Kenya to mainstream indigenous knowledge into tourism as the other socioeconomic development programmes by developing the requisite policies, structures and implementation frameworks. The findings may also be used by tourism sector stakeholders in Kenya to identify, enhance, package and promote indigenous tourism products and services effectively. The findings can also be used by indigenous communities in Kenya to harness the potential of their freely available cultures and physical environment as means of socioeconomic development to complement the other sources of income. The findings may also be used by conservationists to promote community-based responses to the consequences of climate change.

References

1. Adam, H.A.: Recognising natural sacred sites and territories in Kenya: an analysis of how the Kenyan constitution, national and international laws can support the recognition of sacred natural sites and their community governance systems (2012). https://cmsdata.iucn.org/downloads/gaiasns_1.pdf. Accessed 21 June 2014
2. Boven, K., Morohashi, J.: Best practices using indigenous knowledge. Nuffic, The Hague (2002). http://www.unesco.org/most/Bpikpub2.pdf. Accessed 21 March 2015
3. Flavier, J.M., de Jesus, A., Navarro, C.S., Warren, D.M.: The regional program for the promotion of indigenous knowledge in Asia. In: Warren, D.M., Slikkerveer, L.J., Brokensha, D. (eds.) The Cultural Dimension of Development: Indigenous Knowledge Systems, pp. 479–487. Intermediate Technology Publications, London (1995)
4. Fleer, M.: Children's alternative views: alternative to what? Int. J. Sci. Educ. **21**(2), 119–135 (1999)
5. Gachie, J. (n.d.): Cultural tourism in Kenya. http://softkenya.com/tourism/cultural-tourism-in-kenya/. Accessed 22 June 2014
6. Government of Kenya: Cultural tourism resources (2011). http://www.kenyaembassyberlin.de/Cultural-Tourism-Resources.58.0.html. Accessed 22 June 2014

7. Government of Kenya: National tourism strategy 2013–2018 (2013). http://www.tourism.go.ke/ministry.nsf/doc/National%20Tourism%20Srategy%202013_2018.pdf/$file/National%20Tourism%20Srategy%202013_2018.pdf. Accessed 22 June 2014
8. Grenier, L.: Working with Indigenous Knowledge: A Guide for Researchers. IDRC, Canada (1998)
9. Hart, M.A.: Indigenous worldviews, knowledge and research: the development of an indigenous research paradigm. J. Indigenous Voices Soc. Work **1**(1) (2010). http://ahrnets.ca/files/data/3/2011/08/Indigenous%20Worldviews,%20Knowledge,%20and%20Research.PDF. Accessed 17 February 2015
10. Hinch, T., Butler, R.: Indigenous tourism: a common ground for discussion. In: Butler, R.W., Hinch, T. (eds.) Tourism and Indigenous Peoples, pp. 3–19. International Thomson Business Press, London (1996)
11. Kalegi, S.: UNESCO heritage sites in Kenya (2013). http://ecotourismdemystified.com/unesco-world-heritage-sites-in-kenya/. Accessed 22 June 2014
12. Ocholla, D.N., Onyancha, B.O.: The marginalized knowledge: an informetric analysis of indigenous knowledge publications. S. Afr. J. Libr. Inform. Sci. **71**(3), 247–258 (2005)
13. Smith, V.L.: Indigenous tourism: the Four H's. In: Butler, R.W., Hinch, T. (eds.) Tourism and Indigenous Peoples, pp. 283–307. International Thomson Business Press, London (1996)
14. United Nations World Tourism Organization: Tourism Market Trends, 2006th edn. United Nations World Tourism Organization, Madrid (2006)
15. Valle, E., Yobesia, M.N.: Economic contribution of tourism in Kenya. Tourism Anal. **14**, 401–414 (2009). http://www.uib.cat/depart/deaweb/webpersonal/hdeeevv0/archivos/archivos/articulo7.pdf. Accessed 17 February 2015
16. Warren, D.M.: Using Indigenous Knowledge for Agricultural Development. World Bank Discussion Paper 127. World Bank, Washington, D.C. (1991)
17. World Bank: Indigenous Knowledge: Local Pathways to Global Development. World Bank, Washington (2004)
18. Zeppel, H.: Indigenous Ecotourism: Sustainable Development and Management. CAB International, Oxfordshire (2006)

Reintegration of Slovak Researchers Returning to Slovak Companies

Veronika Šramová[1(✉)], Anna Závodská[1], and Viliam Lendel[2]

[1] University Science Park, University of Žilina, Univerzitná 1, 010 26 Žilina, Slovak Republic
{veronika.sramova, anna.zavodska}@uvp.uniza.sk
[2] Faculty of Management Science and Informatics, University of Žilina, Univerzitná 1, 010 26 Žilina, Slovak Republic
viliam.lendel@fri.uniza.sk

Abstract. This paper focuses on understanding the problem of brain drain and identifies crucial activities which must be done in order to successfully reintegrate Slovak researchers. This reintegration process should lead to valuable knowledge sharing, R&D development, educating next generation of excellent researchers as well as increased economic growth. R&D does not relate only to universities but also companies which want to be innovative. Therefore, we focus our research on companies that has R&D department and may be willing to reintegrate returnees working at any research institutions or universities. Based on the knowledge gained from our research conducted in research institutions in Slovakia and best practices from Israel we proposed recommendations (on national, regional and company level) to create systematic approach to reintegration process. These recommendations can significantly affect attitude of researchers towards moving back to Slovakia and bringing the potential value associated with their international mobility to companies.

Keywords: Reintegration · Slovak researchers · Return · Factors · Mobility · Company

1 Introduction

In the literature, one can find a number of theories that try to explain why reintegration of researchers occurs and which factors influence this process. Various authors look at the reintegration process from different angles; however, there is an integrating element to create a comprehensive guideline for the reintegration activities.

In many cases, this effort leads to an excessive complexity and consequently lack of clarity, which often influences the decision of a manager to accept/not accept the researcher from abroad. In many cases a failure in realized reintegration processes occurs and the result does not match the original reintegration objectives of the company. Therefore, an important prerequisite for the success of any adaptation process of researchers working abroad is their smooth integration into company's research team.

Based on the interviews we conducted with a number of relevant institutions within our ongoing project of reintegration of Slovak researchers back to Slovakia a

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 353–364, 2015.
DOI: 10.1007/978-3-319-21009-4_27

requirement to identify the main problems in the process of adaptation of Slovak researchers after their return to Slovakia and proposal of appropriate recommendations that would be beneficial for these institutions emerged.

Our first paper [1] dealt with the reintegration of researchers at the university. To sum it up, it is known that for the last 10 years there have been at least 271 doctorate holders from Slovakia (27 % of them are women) in the field of natural, engineering and technical science. In employment, there are prevailing research workers (229), of which the majority consists of science-education employees and teachers at universities (115), researchers and experts in physical and chemical science, mathematics and informatics (51), researchers and experts in biological science and related science (33) and others. However, it cannot be found out how many researchers have returned after their stay abroad. Furthermore, there is no available national or regional statistics which would monitor this issue.

In view of the seriousness of the current trend of brain drain, we decided to investigate issues of reintegration of Slovak researchers back into Slovak companies. Based on our research we will propose recommendations how to bring back researchers to Slovak companies and how government and self-governing regions can be helpful in this process.

The paper is structured in the following way. It begins with the examination of the current state of international mobility and reintegration of researchers already published in the literature. Next section describes the objectives and methodology of the research. Subsequent section discusses approach to reintegration activities of researchers in Slovak companies. This is followed by highlighting challenges in reintegration of researchers and there are pointed out some recommendations how to face these challenges. The last section concludes the most important ideas from this paper.

2 International Mobility of Researchers

The researchers are a large group of people which is responsible for the development of the country, innovation, revolutionary changes, economic development, creating new knowledge, its sharing, etc. One way how to develop this knowledge and increase its value is to support the research mobility through different countries. Knowledge sharing among researchers can significantly contribute to the development of several countries by increasing "gross domestic product (GDP), gross national product (GNP), the per capita income, level of industrialization, amount of widespread infrastructure and general standard of living" [2]. Knowledge creation, circulation and exploitation are the key elements of modern research and development (R&D) and innovation systems and underpin the evolution of so-called knowledge-based economies and societies [3]. In this sense it is a good sign that "geographical and sectoral mobility is becoming a natural part of the researchers' life" [4].

Fernández-Zubieta and Guy underlined several benefits of the research mobility like "the creation of dynamics networks, improved scientific performance, improved knowledge and technology transfer, improved productivity and ultimately enhanced economic and social welfare" [3]. According to Edler et al. international mobility can be potential option for scientists to improve their access to colleagues, knowledge and skills [5]. The international mobility of researchers has been acknowledged as a powerful mechanism for knowledge transfer (Adams 1968) [6]. The Australian researchers' mobility portal considers "the mobility of researchers to and from Australia important for the vitality of Australia as a knowledge-driven economy within the global community. A mobile workforce is particularly important in promoting knowledge flows and ensuring a diverse and highly skilled workforce that has the capacity to respond to emerging opportunities and challenges [7]".

Research mobility is associated with the famous problem of the brain drain. Cañibano et al. argue about the central phenomenon of the initial debate so-called "brain drain/brain gain" (Adams 1968). However, the inherent nomadism and networking that characterizes the scientific profession, made some scholars reconsider the drain/gain approach and start talking rather about "brain circulation" (Meyer 2001) [6].

A number of factors cause that lots of valuable and clever researchers are moving to another country, e.g. an oppressive political system, lack of job opportunities and social problems such as crime, insufficient health care and low quality education [8, 9], better economic conditions in the field of research and education, scientific and academic environment, research institutes, prestige and access to knowledge, personal relationships, personal and family problems, gaining knowledge and recognition necessary for career advancement in the country of origin etc. [10, 11]. Seventh framework programme (FP7) funded the MOREBRAIN Project (Brain Circulation from Brain Drain to Brain Gain), which was concerned to understand the factors that influence researchers' mobility. The consortium engaged in research "to find out the reasons why researchers pursue a research career out of Europe and remain there and why researchers choose to return to their home country in Europe [12]".

If the governments, companies and research institutions want to enhance their development they need to learn how to use all available means for the support of the brain circulation through the countries and how to transform the researchers' knowledge to added value of final products of their activities (e.g. new knowledge, goods, new regulation, etc.). It requires better facilitation of the research mobility and elimination of the most barriers that prevent researchers' from reintegration. Several organizations, networks and portals are known because they want to help researchers with their mobility. The Euraxess is "unique initiative providing access to a complete range of information and support services for European and non-European researchers wishing to pursue research careers in Europe" [13]. Marie Skłodowska-Curie actions (MSCA) support research training and career development focused on innovation skills in the period to 2020 [14]. The Research Mobility Programme supports the strengthening of "the cooperation and the development of joint research activities in order to broaden the horizons of researchers, to facilitate transnational research, to promote knowledge flows between institutions which promote new interdisciplinary academic networks, transnational training and research mobility schemes within Europe [15]". Each country has own national programmes and initiatives to support research mobility, e.g. the Academy of Finland supports "the international networking and interaction among researchers as well as their personal mobility at different stages of the research career [16]".

The literature provides many views on benefits of the research mobility as well as pull-and-push factors concerning researchers' migration but it still lacks a comprehensive overview of the steps and the guidelines to facilitate successful researcher's reintegration. Therefore we focused on this issue in our research.

3 Research Approach and Methodology

This paper focuses on developing several ways how to facilitate the reintegration process of Slovak researchers when they are planning to return to Slovak companies. It was important to become familiar with the researchers' knowledge about attitudes, benefits, values and barriers which are connected with research mobility and reintegration. The results gained from the realized research have a crucial role in the identification of the main barriers in the companies' reintegration process. A presented case is an example of good practice of the reintegration process in Israel. This case was chosen like basic feasible solution in the comparison with other solutions from the other countries. The proposals for the elimination of main barriers of the reintegration process and defined recommendations are intended for companies and other research institutions. A valuable guideline should contribute to the more effective reintegration process in companies and to value growth of knowledge sharing through countries.

For the purpose of this research we used quantitative and qualitative research approach. Information was obtained and gathered using the document analysis and questionnaire method. The method of quantitative evaluation was mainly used (application of statistical methods) as well as the comparative method in the stage of information processing.

4 Results of the Empirical Research

Research was conducted by Transport Research Institute, JSC. for the needs of the project in which the authors act as principal investigators. The data was collected from October to November 2014 and 69 enterprises participated in it (69 interviews). The target group consisted of *all enterprises in Slovakia. To determine representative sample a technique of simple random sampling was used*. The largest group in viewpoint of capital structure consists of domestic enterprises (84,38 %), then there are enterprises with mixed ownership (9,38 %) and foreign enterprises (6,25 %). The presented research was conducted in Slovakia in order to obtain an overview of the situation in the Slovak enterprises for subsequent proposal of particular steps that will be taken in the Slovak conditions.

The final number of respondents involved in our research was mainly influenced by the existence of two major factors. The first factor was the short duration of our survey coming from the project's needs. The second factor was the existence of a large group of enterprises whose scope of their business does not require collaboration with a research organization. These companies did not consider it important to participate in this survey. On the other hand, just those respondents who have experience with cooperation with research institutions participated in the survey.

Increased attention has been focused on the issue of adaptation of researchers in Slovak companies after completing their research mobility within the survey *Identification of the level of cooperation of Slovak companies with research organizations.*

The first area of investigation was to find the *current status of hiring of researchers who have worked abroad for some time* by Slovak companies. 17,19 % of Slovak enterprises said that they are currently employing researcher who worked some time abroad. The majority of Slovak companies (76,56 %) currently does not employ such a researcher and 6,25 % of the respondents did not respond to the question.

Further exploration area was *identification of the main difficulties hindering employment of researcher who works/worked for some time abroad.* Most perceived problems are financial costs (45,3 %), pointless employment of such a worker in the particular company (43,8 %), complicated process of finding suitable candidates (25 %) and time consumption (17,2 %). An interesting identified difficulty is the lack of interest in this type of worker from the company's management side (12,5 %).

The main area of investigation was the *area of supporting an adaptation of researcher from abroad into the company*. By supporting the adaptation we mean providing advice, finding a suitable accommodation, explaining researcher his job responsibilities and duties, introducing to his team, building positive relationships between current employees, etc.

Interviewed companies could express their opinion on how important this criterion is in assessing readiness to accept researcher from abroad on a scale from 1 to 5, where 1 is "irrelevant" and 5 "extremely important". They could also give their view on how their company is actually fulfilling this criterion - Ensuring adaptation in their company. Again, they could use 5 point scale, where 1 means "worst performance" and 5 means "best performance".

The majority of Slovak companies (56,25 %) attaches great importance (level 4 and 5) to supporting adaptation. Only 21,88 % of respondents expressed the low importance (level 1 and 2) for this criterion. The average importance (from all numerical values of respondents) was 3,28. When considering the performance of Slovak enterprises in this area it can be concluded that in the majority of enterprises (65,63 %) can be seen significant or less significant points for improvement (level 1, 2 and 3). Conversely excellent performance (level 4 and 5) has only 34,37 % of Slovak enterprises. The average performance in the field of supporting adaptation is at level 3,03. This represents a value lower than the average importance.

5 Israel as a Case of Good Practice

During the implementation of the survey our attempt was to find out how selected countries implement their reintegration policy and what they can offer to their researchers when they have an interest to return to their home country. Based on acquired information we selected countries such as Israel, the Netherlands, Norway, Finland, Germany and the United States because of their economic development, state of science and research, openness to research mobility as well as hiring researchers from other countries. These countries have strong interest in helping researchers in their mobility and subsequent return but this support is not systematic with exact rules, on the contrary it is more intuitive. These countries do not suffer from brain drain in such a great extent; therefore, they are not quite suitable to be used as good practice. However, Israel is dealing with the same problem of brain drain. Their system is well elaborated, supported by government and thus we chose this country's model for our paper's purposes. When looking for solutions in V4 Countries (the Czech Republic, Poland and Hungary), which are historically and economically alike to Slovakia we found out that these countries also have to deal with similar problems of the reintegration. Nevertheless, they do not provide superior solutions than the case study of Israel.

All reintegration processes are supported and incorporated by the Israeli government. "The Ministry of Aliyah and Immigrant Absorption offers new-immigrant and returning-resident scientists and researchers diverse forms of assistance for promoting integration into the R&D sector in Israel [17]". They created The Center for Absorption in Science which "helps immigrant scientists find employment in institutions of higher education, research institutes, technological incubators, and in the industrial and business sectors. The Center recognizes the significant contribution that immigrant scientists make to the Israeli economy and assists them in obtaining suitable conditions for the R&D [18]".

If immigrants and returning resident scientists have interest in helping them with reintegration in Israel The Center for Absorbtion is committed to ensure the following in accordance with the regulations of Israel [18]:

- To facilitate the absorption of the immigrant or returning resident scientist into the R&D system and academic research institutions in Israel.
- To advise and guide scientists looking for work in Israel.
- To help enterprises and academic institutions in the market to absorb immigrant and returning Israeli scientists.
- To encourage the expansion of the R&D system in Israel while using the knowledge and experience that these scientists bring with them.
- To participate with other government facilities in determining the policies concerning scientific personnel development in Israel.

The above case shows that there is a need for strong government support in the process of reintegration, which is currently lacking in Slovakia.

6 Identification of Main Problems and Proposal of Reintegration Activities on National, Regional and Company Level

A survey of foreign literature and examples of good practice from abroad show the importance of various institutions involved in promoting brain circulation and reintegration of researchers as well as the existence of a unified re-integration system, which would be supported by the government of the country. However, the research conducted in Slovak enterprises showed that even enterprises themselves are not systematically prepared for adoption of reintegrated researchers. Therefore, the main identified problems and recommendations are focused on improving the situation directly in enterprises and on national and regional level. We believe that these are the three essential pillars in building the reintegration policy in Slovakia. Recommendations were created in cooperation with a successfully reintegrated researcher. He finished his doctoral studies at Ghen University, Belgium (2001–2005) and conducted his research in the field of Numerical methods for computing electromagnetic fields till 2010. Then he returned with his family to Slovakia, where he continues with his research at the University of Žilina. All his comments based on his experience were taken into account in recommendations.

Recommendations for Enterprises. Adaptation process of reintegrated researchers in the enterprise is a complex process that must be based on a detailed analysis of reintegrated researchers and their needs, plus on the other hand job requirements and working environment of the enterprise itself.

The first problem area in the process of adaptation of reintegrated researchers in enterprise is insufficient development of adaptation program. It does not match all of objectives and tasks of adaptation of reintegrated researchers in enterprise. A common mistake is relying of HR department on the fact that the newly recruited reintegrated researcher will become familiar with the work environment by himself. Poor information and lack of documentation, which is the basis for the proposal of adaptation program, leads to its insufficient elaboration. We recommend HR manager to *carry out a detailed analysis of the working environment, jobs and reintegrated researchers who are subject to a process of adaptation*.

Another identified deficiency is *not engaging reintegrated researchers into making a draft of an adaptation program*. If the adaptation process should be successful, it is necessary that enterprise knows opinions and attitudes of their already employed reintegrated researchers and discuss with them working draft of the adaptation program. We recommend HR manager to create conditions for regular and open communication with reintegrated researchers in the adaptation process. It is also necessary to ensure that an enterprise collects ideas and preferences of newly reintegrated researchers and subsequently takes them into account when creating the adaptation program.

Problems related to the *lack of feedback* may be experienced within the adaptation process of reintegrated researchers. This situation is caused by the fact that the enterprise does not have valuable information based on which it would be able to take the necessary steps to improve the whole adaptation process of reintegrated researchers in the future. The benefits of the process of adaptation cannot be assessed by enterprise because it does not have defined any measurable targets. On the other hand enterprise does not ensure continuous observation of the adaptation process of reintegrated researchers.

HR manager should focus on *defining the objectives of adaptation* (clear, understandable, time-bound, feasible) *and measurable indicators*. Consequently, it is necessary to choose appropriate metrics that the enterprise will use for assessing the fulfillment of the goals of adaptation process. The metrics must be based on the main priorities of the orientation of reintegrated researchers in enterprise. All these activities should lead to ensuring continuous monitoring during the adaptation process of reintegrated researchers which will bring valuable information for the future improvement.

Biggest problem within the adaptation process of reintegrated researchers in enterprise is lack of unified approach that would regulate the process in terms of the division of competencies, its content and evaluation. The procedure will define requirements to maintaining documentation and records of the whole process of adaptation of reintegrated researchers. HR manager should understand the whole adaptation process as a *complex system* that is based on the *identification of adaptation needs, evaluating the effectiveness of the adaptation program and the methods used*. Enterprise should also establish appropriate organizational and institutional preconditions.

Newly recruited reintegrated researchers have problems with orientation in the new enterprise. Getting to know working relationships, employee benefits, principles and values of the enterprise is therefore very essential. These facts influence the adaptation process of reintegrated researchers and slow it down. A suitable solution is the use of written documentation in the adaptation process. However, it is necessary to take into account the character of the job, the level of integration of researcher, his skills and abilities, etc. Effective in this area can be *personnel manual (containing human resource policies relating to research staff) and orientation package*.

Recommendations on National and Regional Level. In order to make the reintegration process more systematic support should come from multiple stakeholders - not just from enterprise but also from government as well as regions of Slovakia.

Recommendations at the state level can be various. Fundamental should be legislative changes that facilitate the reintegration process. Very important is information support in the form of providing current information to researchers through various web portals, which should contain information on the possibilities of reintegration, e.g.:

- The Ministry of Education, Science, Research and Sport of the Slovak Republic: www.minedu.sk,
- The Ministry of Foreign and European Affairs of the Slovak Republic: http://www.foreign.gov.sk/,
- Slovak Research and Development Agency: http://www.apvv.sk/,
- Slovak academic information agency: www.saia.sk,
- EURAXESS in Slovakia: http://www.euraxess.sk/sk/,
- Academic Ranking and Rating Agency (ARRA): www.arra.sk,
- and other institution related to research and mobilities.

Additional information tools could be social networks where groups that are focused on the issue of reintegration could be created. Through discussion forums researchers could be allowed to share information and advice to each other. Current information could be sent in newsletters, which could be distributed to individuals and institutions involved in the reintegration problematic. Newsletters could include the following:

- information on the funding of science in Slovakia,
- examples of good practice on the reintegration of researchers,
- interviews with relevant persons,
- job offers from institutions that are interested in Slovak researchers working abroad,
- articles addressing serious issues related to reintegration.

The above mentioned activities should be the means for creating communities (institutions and people) that will address particular issues and deal with the situation of reintegration of researchers.

Within the region, the information can be spread through available information resources that regional governments have. Provided information should be targeted and focused on addressing reintegration issue within the region. Other information sources can be local print and broadcast media, which can address the issue of reintegration of researchers in accordance with their focus.

At the regional level following institution should be involved in the reintegration issues: self-governing regions, cities, universities and competence centers for reintegration.

Within self-governing regions reintegration could affect areas such as: transport and regional development, information and foreign relations, education and sports. It is also important to involve institutions that are focused on promoting regional development and transfer of R&D results into practice, such as universities and business incubators.

Mentioned institutions should jointly focus on finding specific solutions that will lead to creation of reintegration system at the regional level. Fundamental is to correct deficiencies and the main barriers to reintegration of researchers who are interested in reintegration (we proposed recommendations for stimulation of researchers to come back to Slovakia in our previous paper [1]).

Slovakia also with the support of the government and EU funds is currently experiencing big boom in the development of scientific research environment by developing science parks and research centers, procuring high technology, increased support for the cooperation between scientific institutions and enterprises, which lead to the creation of space where research workers can return from abroad. One of the biggest problems, however, remains an adequate remuneration of new staff and experts returning from abroad.

The solution can be sharing costs on researcher by financing him from several sources, e.g. from programs financed by the EU, regional authorities and universities or companies where the researcher works. If the researcher is starting his new project, which is innovative and have commercial potential, he could be placed in one of the incubators and accelerators in Slovakia. In this case researcher will be funded directly from the resources of the accelerator for a certain period. Another solution can be a *remuneration package* for enterprises which employ a researcher (e.g. tax allowance, shared funding of the researcher' s salary, etc.).

In addition to financial issues, researchers must also deal with housing issue. Institutions where researcher supposes to work should either provide him accommodation until he finds something suitable or directly help him find appropriate accommodation.

It is necessary to realize that researcher has a family that would like to come back to Slovakia with him. Therefore, institutions should help researchers find suitable jobs for their partners as well. If researchers have children, it is necessary for them to find nursery, kindergarten, school, etc. In this case the state support of English kindergartens is appropriate (when partner is not Slovak children may not speak Slovak).

A key factor in the development of reintegration support in Slovakia could be the establishment of reintegration centers at regional level. These centers could, in cooperation with institutions (universities or companies), help to solve all of the above mentioned requirements that reintegrated researcher must arrange before he starts to work at new research facility in Slovakia. This institution should help researchers to address all reintegration issues so the researcher obtains all necessary information, go through all legal and bureaucratic procedures that are necessary for reintegration. Institution could also be able to order such services from the reintegration center which would oversee the whole process of reintegration. The previous research confirmed [1] that mobility of researchers is not thoroughly monitored and therefore it is very difficult to gain any information about researchers and their stays abroad. Reintegration centers can start to monitor researchers' migration. Only in this way it will be possible to communicate with researchers and solve their problems.

7 Conclusion

Many countries have found that they can gain many positive effects through sharing of knowledge in the brain circulation process and therefore they began to promote it. Developed countries possess first-class technology and thanks to their educational system and excellent R&D do not face the problem of brain drain. For knowledge sharing they use various other means, e.g. IT tools, conferences or they simply invite those researchers who could be beneficial for their institution. Regarding mobility they generally have no problem with the fact that their researchers do not want to return back to their home workplace and if it happens they have no problem finding their replacement.

On the contrary, countries that face the problem of brain drain and identified barriers for reintegration must find ways how to achieve that researchers after their stay abroad came back to their home country - whether to companies or other research organizations.

The reason for this effort is the fact that the less successful countries require active participation on brain circulation and know how to use the value of knowledge capitalization on their domestic workplace.

In our new research, we moved few steps further and on the basis of our proposals tailored to the business environment, we can continue to work on developing a systematic basis for effective and targeted reintegration of researchers. Bolder goal in the future is that this strategy will be introduced to research institutions and government representatives who should actively participate in shaping reintegration policy in Slovakia and exploit benefits of brain circulation. We proposed recommendations on national and regional level and we hope that they can be a crucial help for responsible persons in creation of complex reintegration policy in Slovakia.

Implementation of recommendations will have positive impacts on reducing the level of unemployment rate in several regions in Slovakia and increasing the number of researchers working in applied research. They will create an active interconnection between research and business environment and improve knowledge sharing and technology transfer. It will enhance the productivity in innovation development, enterprises' turnover and their competitiveness which will result in higher number of researchers.

Acknowledgements. This paper is published as one of the scientific outputs of the project "University Science Park of the University of Žilina ITMS: 26220220184 supported by the R&D Operational Program funded by the European Regional Development Fund".

References

1. Šramová, V., Závodská, A., Lendel, V.: Identification of motivational factors influencing the return of researchers focusing on the Slovak environment. In: Uden, L., Oshee, D.F., Ting, I.-H., Liberona, D. (eds.) KMO 2014. LNBIP, vol. 185, pp. 56–66. Springer, Heidelberg (2014)
2. Investopedia 2015: Developed Economy. In: Investopedia, Dictionary (2015). http://www.investopedia.com/terms/d/developed-economy.asp#axzzlegO8olO
3. Fernández-Zubieta, A., Guy, K.: Developing the european research area: improving knowledge flows via researcher mobility. In: JRC Scientific and Technical reports, European Commission (2010). http://library.certh.gr/libfiles/PDF/MOBIL-103-DEVELOPING-ERA-by-ZUBIETA-in-JRC58917-Y-2010.pdf
4. The Faculty of Mechanical Engineering at the University of Zilina: Projekt EURAXESS - Mobilita výskumných pracovníkov (2014) http://www.fstroj.uniza.sk/projekt_euraxess_-_mobilita_vyskumnych_pracovnikov
5. Edler, J., Fier, H., Grimpe, C.: International scientist mobility and the locus of knowledge and technology transfer. Res. Policy **40**, 791–805 (2011). http://www.sciencedirect.com/science/article/pii/S0048733311000382
6. Cañibano, C., Otamendi, F.J., Solís, F.: International temporary mobility of researchers a cross discipline study. Scientometrics **89**(2), 653–675 (2011). http://link.springer.com/article/10.1007/s11192-011-0462-2#page-1
7. Australian Researcher's Mobility Portal: Are you a researcher planning to move to, or from, Australia? (2014) http://mobility.anu.edu.au/
8. Thorn, K., Holm-Nielsen, L.B.: International Mobility of Researchers and Scientists. Policy Options for Turning a Drain into a Gain. UNU World Institute for Development Economics Research. Helsinki. Research Paper No. 2006/83, (2006) http://www.econstor.eu/bitstream/10419/63363/1/51649791X.pdf
9. Bhagwati, J., Hamada, K.: The brain drain, international integration of markets for professionals and unemployment: a theoretical analysis. J. Dev. Econ. **1**(1), 19–42 (1974)
10. Baruffaldi, S.H., Landoni, P.: Return mobility and scientific productivity of researchers working abroad: the role of home country linkages. Res. Policy **41**, 1655–1665 (2012)
11. Jonkers, K., Cruz-Castro, L.: Research upon return: the effect of international mobility on scientific ties, production and impact. Res. Policy **42**, 1366–1377 (2013)
12. Marie Curie Fellows Asiciation "Research without frontiers: MOREBRAIN project: understanding the factors that influence researchers' mobility (2011) http://mcfa.eu/site3/?q=node/13
13. European Commission: EURAXESS - researchers in motion. In: The EURAXESS Jobs Portal (2014) http://ec.europa.eu/euraxess/index.cfm/general/about
14. European Commission: Marie skłodowska-curie actions – overview. In: European Commission, Research and Innovations, Marie Skłodowska-Curie actions (2014). http://ec.europa.eu/research/mariecurieactions/about-msca/actions/index_en.htm
15. Internet – Science: Researcher mobility. In: Network of excellence in internet science (2014) http://www.internet-science.eu/mobility

16. Academy of Finland: Researcher mobility. In: Funding and Guidance, The research career (2014) http://www.aka.fi/en-GB/A/Funding-and-guidance/The-research-career/Researcher-mobility/
17. Ministry of Aliyah and Immigrant Absorption: Scientists and researchers. In: Ministry's Website, Returning to Israel (2014) http://www.moia.gov.il/English/RETURNING RESIDENTS/Pages/MadaUmechkar.aspx
18. Ministry of Aliyah and Immigrant Absorption: Center for absorption in science. In: Ministry's Website, ministry Units (2014) http://www.moia.gov.il/English/About/Office Units/Pages/kliteBemada.aspx

Knowledge Management Platforms and Tools

Possibilities of Streamlining Within Business Intelligence Systems in Business Practice

Milan Kubina[1], Gabriel Koman[1], Michal Varmus[1(✉)], and Luboš Takáč[2]

[1] Department of Management Theories, Faculty of Management Science and Informatics, University of Žilina, Univerzitná 1, 010 26 Žilina, Slovak Republic
{milan.kubina, gabriel.koman, michal.varmus}@fri.uniza.sk
[2] University Science Park, University of Žilina, Univerzitná 1, 010 26 Žilina, Slovak Republic
lubos.takac@uvp.uniza.sk

Abstract. Management and decision-making in modern companies currently significantly affect information collected and distributed to managers through a variety of information and communication technology (ICT). The amount of information is growing due to modern technologies, their constant development and improvement in companies. This large amount of information companies has to be processed and therefore complex information systems such as Business Intelligence system are used more often. These systems are designed to support deciding of leading workers in the company. They are very difficult and expensive with regard to information infrastructure, technical equipment and staff of the company. To ensure efficient use of these expensive systems it is essential to provide constant mapping and implementing the new trends in the development and use of the systems. This paper describes the possibility of improving efficiency within business intelligence systems in companies.

Keywords: Business intelligence · Data · Management · Information system · Technologies

1 Introduction

The influence of new technologies and techniques in the area of business processes, such as production, finances, marketing, etc., there constantly arises in the companies larger and larger amounts of data. These data are important for the company and are also entry into the decision-making process. Business Intelligence Systems are used in companies as comprehensive decision support systems of managers. System allows analyzing data stored in a single database structure and generating reports with relevant information value for leading company managers. Business Intelligence provides the collection, sorting, distribution and presentation of relevant information for the needs of strategic planning and business management in the company. Business Intelligence (BI) systems are developing together with ICT, new trends and possibilities for streamlining of these systems in the form of various procedures, models, applications, integration are created.

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 367–376, 2015.
DOI: 10.1007/978-3-319-21009-4_28

It is important to search solutions for streamlining, either by updating software equipment, staff training, system integration with company systems or its implementation in the company, etc. to ensure using of the system in the future and also return on investment. By searching and implementing of new opportunities for improving the system efficiency, companies can achieve reduction of operating costs on the system, downtime elimination, faster return on investment, profit increase, discovery of new business opportunities based on quality and quickly available information, achievement a high employee qualification level for work with the system, and many other benefits.

2 Business Intelligence

Nowadays, the Business Intelligence systems are getting more often part of the information system, especially of large companies but also of the medium-sized ones. Development of these systems was particularly affected by the significant leap in the development of information technology and the continuing growth of competition. The development of the information technologies (IT) is associated with generating of big data amount, which analysts must process and offer managers for decision-making.

By using of available technology it was possible to process this amounts of information effectively, easily and quickly make available and show opportunities and danger situations for the company. From this reason there were created new departments in the form of data warehouses, Business Intelligence systems and many more. Business Intelligence is defined as systems that collect, transform, and present structured data from multiple sources (Negash 2004) reducing the needed time to obtain relevant business information and enable their efficient use in management decision making process (Den Hamer 2005), allowing dynamic enterprise data search, retrieval, analysis, and explanation of the needs of managerial decisions (Nofal et al. 2013). Pirttimäki (2007) describes Business Intelligence as a process that includes a series of systematic activities, being driven by the specific information needs of decision makers and the objective of achieving competitive advantage. According to Tyson (1986), Business Intelligence focuses on collecting, process and present data concerning customers, competitors, the markets, technology, product and the environment.

Principle of data processing in BI systems is based on secondary data, which is different from what the transaction systems which work with - primary data ("single-use data"). By secondary data processing, the data are used in a different way than they were initially generated for. It comes to information linking that arises within different activities and time. The information is then analyzed and used in the optimization, management and decision-making. Secondary systems read data generated in large amounts, using advantage of tricky methods of analysis, aggregation, statistics, etc. These IT systems are more difficult in terms of performance than the primary ones. The main advantage of these systems is that they analyze the data generated in different time intervals and different systems.

Contents of the BI concept are the skills, knowledge, technological equipment, application equipment, security issues, models and techniques used in business for the purposes of analyzing and understanding the market situation. The purpose of BI

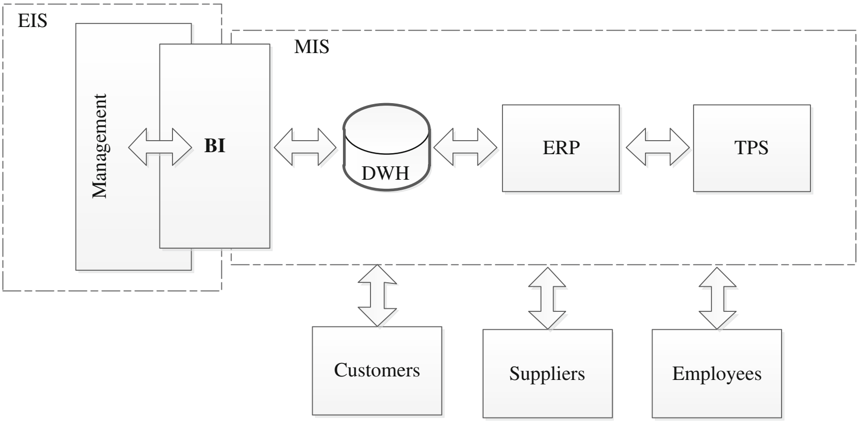


Fig. 1. Position of BI in information system of the company

is therefore to support business decisions. Applications run within the BI system allow analysts to work with historical, current and predicted business operations and corresponding data using historical data warehouse or data direct from operational systems (Carnicky 2006, Novotny 2005).

Basic features of BI include: Online Analytical Processing (OLAP), Reporting, Support analysis, Overview mode (Dashboard, Balanced Scorecard), Data mining, Corporate Performance Management (CPM), Predictive analysis (Jenco 2011).

Among forms of user presentation of outcomes BI includes, in particular: assemblies, Questions, OLAP, control panels, summary of the results (Novotny et al. 2005, Bouman and Dongen 2009).

Within the BI system there are used various specific analysis that use a lot of information useful for management and decision-making. Therefore BI systems are not focused on basic processing and implementation of current transactions (production, trade or finances). The role of BI is the support of senior management decision-making processes by providing relevant and necessary information. That is why its position is above the rest of the information systems (Fig. 1).

3 Use of Business Intelligence in Companies

Business Intelligence systems belong to the information systems that are currently used mainly in large companies for analysis of data amounts. These data are generated by operating systems of the company in combination with other data sources. Source data contain important information that business analysts obtain by their analysis.

While Business Intelligence is a comprehensive information system, analysts keep at disposal a variety of automated and analytical tools and functions necessary for data processing. Among these tools the most common are the following: ETL (Extract, Transform and Load), OLAP, data mining, reporting, etc. (Hamranova 2013). Business Intelligence systems are also suitable for small and medium-sized companies generating large volumes of data. There exists use of BI systems, particularly within the automation of routine business processes (production, trade, finance) in these companies (Carnicky 2006). The information obtained by analyzing data from various data sources and time intervals, serves as inputs to support activities in the management and decision-making company. Business Intelligence systems are often used in all business processes in the business practice.

PARAMIT company uses Business Intelligence system for formation of time-based analysis, for obtaining relevant information about the needs of customers, or inventory management (Solutions for managerial decisions in the ERP Money S3 2014). Association ISIC Global Office gained ability effectively by the implementation of Business Intelligence system gained manage the flows of data for information support within strategic decision-making. The system allowed the association effectively interconnect, save and distribute data from different data sources and those processed through advanced reporting and analytical tools with a simple user interface (Business Intelligence for ISIC Global Office 2012). For summary of customer records and their orders MOL Slovakia is also using Business Intelligence systems. The system allowed the unification of all customer processes and integration of all systems generating data that are based on the Microsoft platform (Integrated solution Microsoft Dynamics NAV and CRM supports key processes in company MOL Slovakia 2010).

ZSSK, Inc. uses the Business Intelligence system within large companies in the Slovak market. The supplier of the system for ZSSK is SAP Slovakia. Business Intelligence system consists of SAP ERP system modules of the, specifically of the modules FI (Financial Accounting) and CO (Controlling). Additional system components include Business Intelligence SAP BW (data warehouse) as the basic structure of the database and SAP BO (Business Object) as the basic system for the integration of all system components as well as the presentation layer on the data warehouse. Thanks to Business Intelligence system, the company has an overview of all data in the data warehouse, which analysts can analyze through OLAP technology and then create and distribute reports to other users of the system for processing or managers for decision making.

As the Business Intelligence system in ZSSK only standard reports are created and processed, the potential of system is under-utilized. Inefficient use of the system makes the fact that the systems used within the section controlling and strategies for the department. Other departments of the company have access to the system in its operations (ZSSK 2014).

While the Business Intelligence systems are very complicated and expensive, not just in technology that is available in company but also level of education and qualification of employees. For this reason, these systems are mostly used by medium-sized and large companies. It is important that a company that wants to implement a Business Intelligence system realized if it really needs this system and will use it effectively. Another important fact is the constant development of ICT, what means that companies need to plan for the future, not only updating the system itself, but also the entire IT infrastructure. Otherwise, the implementation of Business Intelligence system will be inefficient investment for the company is in the future.

4 Suggestions for Improving Efficiency in the BI Companies

For constant improvement of the efficiency of Business Intelligence system in the company it is necessary that the system is regularly updated, is used throughout the company and that qualified staff knows how to work with it. Improving performance can thus be achieved, for example: by using of new technologies in the embedded system software support, by implementation of the system within the company, by direct integration with company systems running Business Intelligence, providing qualified staff to work with the system.

4.1 Use of New Technologies

Constant development of modern technologies in hardware and software is gradually emerging new trends within the processing and presentation of data. For increase of the system efficiency, it is appropriate to consider the establishment of businesses and other trendy software solutions in the processing and presentation of data. In this way it is possible to simplify and speed up the work of employees in creating graphical reports mainly in the form of various graphs, based on the selected attributes. Another benefit of the new software solutions is linking the system with mobile applications for presentation. While the mobile market constantly grows and develops, there is a presumption that these devices will be in the future, the underlying hardware used in presentations, checks of the company status, or these devices will be used within the decision-making of managers.

As an example of new software solutions we can include SAP Lumira for Business Intelligence systems based on SAP platform. The SAP Lumira can easily and quickly perform analysis of data from different sources and of different sizes and combinations (see Fig. 2) (SAP Lumira Standard Edition 2015).

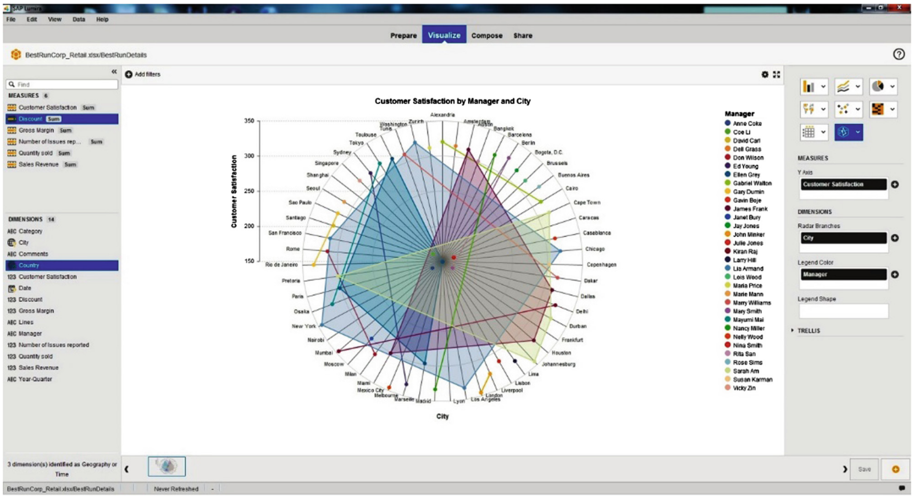


Fig. 2. Graphical report of the SAP Lumira program on the base of OLAP analysis

It is possible to retrieve data from Microsoft Excel into application, or directly from relational databases. This analytical tool allows through point-and-click visualizations in drag-and-drop interface to create different graphical evaluation of the retrieved data based on OLAP analysis. Analysts have within SAP Lumira available also SAP LumiraCloud, which is also connected with a mobile solution SAP BI, what allows managers quickly and easily access data and reports generated in the cloud. Using of new software solutions within the Business Intelligence systems work of employees can be accelerated and streamlined because they have the ability to create reports quickly and easily from various data sources and to quickly share these to the cloud for the needs of managers.

At the same time the quality of decision-making processes of managers will be improved because they will have the necessary data to which they have unrestricted access also via mobile devices.

4.2 Integration of System Business Intelligence with Mobile Devices

By impact of new technologies, mobile devices like tablets, smartphones, etc. are useful in business processes. Mobility is continually increasing trend in the IT field, what shows the sale of mobile devices (smart phones), which increased by almost 40 % in 2012 compared to 2011. Up to 72 % mobile market uses the Android platform (Google) and the other 13.9 % iOS platform (Apple) (Surveys–Results from surveys about the Slovak IT market 2013).

Therefore is integration of Business Intelligence system in companies with mobile devices important, in terms of improving the efficiency of the system and its use in the future.

Linking with BI company is based on the company-based user authentication data (username and password), which connects through Internet to BI system in company, or the cloud, where there are accessed and shared data (databases, reports and so on.) (SAP BusinessObject Explorer 2013, 2015).

The results of the analysis can be shared further through mobile applications with another manager for the needs of decision-making or cooperation on given issue.

Users can use for analysis in applications similar graphics as in the BI system (Fig. 3), that means to create clear graphical reports that can be changed in real time by changing the source data.

4.3 The Implementation of the System in All Parts of the Company

Business Intelligence System is a comprehensive information system for decision support with lots of analytic functions and tools. From this reason it is appropriate that this system was implemented across the company. Using the system only for selected areas of business, such as in the case of ZSSK, Inc. is inefficient. By the implementation of the system to other areas of the company it is possible to ensure the creation of a new standard for data processing and sharing the resulting reports. At the same time there will be unified system with a single service, training and qualifications of staff in

Fig. 3. Working area of BI (SAP) on the Android platform

these areas of the company. Thanks the implementation of BI system to all areas of the company, it is also possible to reduce the burden on departments that provide the company processing of outputs from other systems. By the implementation of the system in the company it is possible to improve ultimately the effectiveness of BI in the company.

Optimal solution Business Intelligence system implementation is shown on the Fig. 4, on the organizational structure ZSSK, Inc. (ZSSK 2015).

By the implementation of BI to all areas of the company data will be used in database system structure not only in selected areas of the company, but also in other areas of optimization for the needs of the present, modeling of future plans, prospects and business simulations.

4.4 Direct Company Systems Integration with Business Intelligence

Most of large and medium-sized companies currently use a variety of information systems for the needs of their business. Data generated by these systems are there processed or they are processed by another system. If Business Intelligence system is implemented in such defined ICT structure of the company, it is important for the efficiency to ensure that data are forwarded from the systems into a unified database structure BI system, before it is processed in the source systems. In case that adaptation data are forwarded to the BI system, they can cause some distortion or damage of the informational value of the original, root data. Since the purpose of the system is to detect opportunities and threats specially from the source data, this method of data distribution is inefficient, because information value of the original data is lost. It is

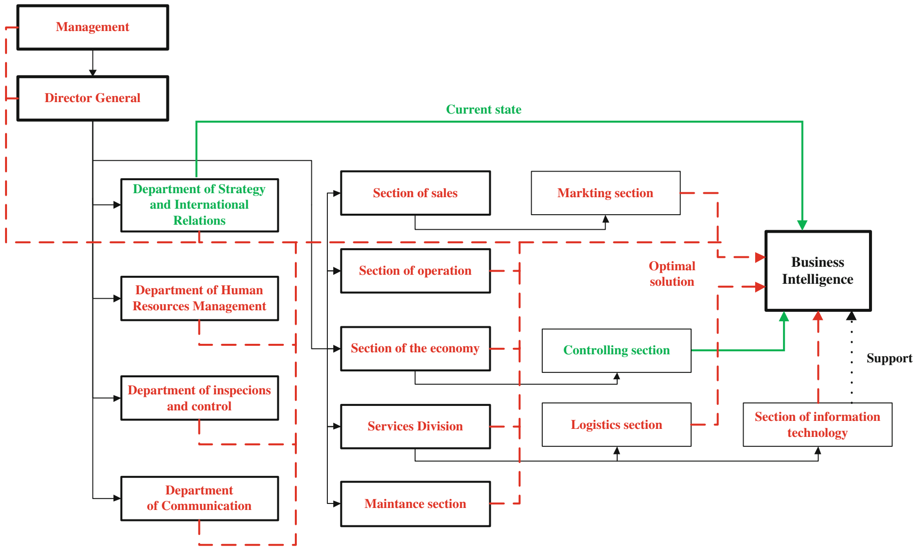


Fig. 4. Optimum state system of the implementation of the system in all areas of ZSSK

therefore necessary that all systems of company be directly integrated with BI system. Integration will ensure forwarding of raw data generated by individual systems in the company, into the database structure on which the BI system is based (most often it is a data warehouse). By system integrating, data will be stored in one place, which will ultimately speed up the transfer of data from individual systems for the needs of primary analysis and the comparison needs further adjusted data from different systems with original data. At the same time, all data stored in one place, in one database structure and whole staff will have access to them with access to the data warehouse. By integration of systems it is possible to eliminate downtime caused by waiting of analysts for processed data from other systems. Data warehouse will fulfill its main task what is to store historical data. This allows using the system more effectively.

4.5 Ensuring of Qualified Staff

In order to streamline the system BI system in the company, it is necessary that the system is used. Essential prerequisite is that the company had employees who know the system and know how to work with it. From this reason trainings should be carried out, aimed at familiarization with system, its work, administration, new trends and technologies in data processing and analyzing.

Employees will be those who will work with the system, therefore it is necessary to provide trainings focused on operation and use of BI system. Through trainings employees will receive the necessary qualifications for the system and an overview of the BI system and its application in the company. At the same time employees will get knowledge of new trends in data processing, or the new software solutions within the BI company. At the same time training will help to eliminate downtime in the form of support calls when working with the system, what will increase the overall effectiveness of all processes within the BI system, work of employees will speed up and use of BI system will increase as well.

Training will enable enterprises to obtain qualified staff, which will be able to work effectively with the system, thus improving their performance, which is ultimately reflected on the efficiency and flexibility of the whole company. At the same time downtime will be eliminated in the form of technical support to employees who need it in their work because they can't work with the system.

5 Conclusions

Implemented Business Intelligence system has great potential for the processing of different types of data from various sources, from which it is possible to obtain a variety of information, knowledge, or even reveal competitive advantage. It is therefore necessary for companies to know how to use it and with its help not only analyze and reveal business opportunities, but also increase employee performance, simplify their work, improve the decision-making process of managers through quickly available relevant data, or remove the amount of downtime due to inefficiency of the existing IT system of the company. It is highly fiduciary and cost ineffective, if the companies own information system, which is not used, developed and updated according to the latest technology trends. If misused, such system can cause companies existential problems.

The most important element in the use of each system is the man who this system operates. That is why the professionalism of staff, acquired by a suitable integrated retraining in the use of Business Intelligence system in their work is necessary. Thanks to qualified employees, it is possible to increase the effectiveness of Business Intelligence system, because they will understand this system, they will be able to work and bring new ideas for its development within the whole company.

Acknowledgement. This paper is supported by the following project: University Science Park of the University of Zilina (ITMS: 26220220184) supported by the Research and Development Operational Program funded by the European Regional Development Fund. KEGA 035ŽU-4/2013.

References

Bouman, R., Dongen, J.: Pentaho Solutions: Business Intelligence and Data Warehousing with Pentaho and MySQL. Wiley Publishing, Indianapolis (2009)

Business Intelligence for ISIC Global Office (2012). http://www.orchitech.cz/documents/10198/13102/Case-Study-ISIC-Business-Intelligence-CZ.pdf. Accessed 7 January 2015

Carnicky, S.: Business intelligence in managing companies and approaches to its solution. Bus. Rev. **5**(9), 36–52 (2006). ISSN: 1335-9746

Den Hamer, P.: The organisation of Business Intelligence. SDU Publishers, The Hague (2005)

Hamranova, A.: Aspects of implementing Business Intelligence to Slovak companies. Economist, Bratislava (2013). ISBN 978-80-225-3603-5

Jenco, M.: Information systems in Organisation. VERBUN, Ruzomberok (2011). ISBN 978-80-8084-780-7

Negash, S.: Business intelligence. Commun. Assoc. Inf. Syst. **13**(1), 177–195 (2004). ISSN: 1529-3181

Nofal, I.M., Yusof, Z.M.: Integration of business intelligence and enterprise resource planning within organizations. Procedia Technol. **11**, 658–665 (2013). doi:10.1016/j.protcy.2013.12.242

Novotny, O., et al.: Business Intelligence. Grada Publishing, Praha (2005). ISBN 80-247-1094-3

Pirttimäki, V.H.: Conceptual analysis of business intelligence. S. Afr. J. Inf. Manage. **9**(2), 1–17 (2007). ISSN: 2078-1865

Solutions for managerial decisions in the ERP Money S3 (2014). http://www.money.sk/money-s5/odvetvove-riesenia/business-intelligence/. Accesssed 6 January 2015

Integrated solution Microsoft Dynamics NAV and CRM supports key processes in company MOL Slovakia (2010). http://www.microsoft.com/slovakia/studie/mol.aspx. Accesssed 7 January 2015

SAP Lumira Standard Edition (2015). http://play.google.com/store/apps/details?id=com.sap.mobi. Accesssed 19 February 2015

Surveys–Results from surveys about the Slovak IT market (2013). http://static.itnews.sk/a542/file/item/sk/0000/07_prieskumy.C6iy.pdf. Accesssed 8 January 2015

SAP BusinessObject Explorer (2013). http://itunes.apple.com/us/app/sap-businessobjects-explorer/id422267790?mt=8. Accesssed 20 January 2015

SAP BusinessObjects Mobile (2015). http://play.google.com/store/apps/details?id=com.sap.mobi. Accesssed 19 February 2015

Tyson, K.W.M.: Business Intelligence: Putting It All Together. Leading Edge Publications, Lombard (1986)

ZSSK (2015). http://www.slovakrail.sk. Accesssed 27 January 2015

ZSSK's internal documents (2014). Accesssed 12 December 2014

Framework for Building a Big Data Platform for Publishing Industry

Aravind Kumaresan(✉)

University of Vaasa, Wolffintie 34, 65200 Vaasa, Finland
aravind.kumaresan@gmail.com

Abstract. The word Big Data is commonly used and it is not new today. Large, medium and small companies are starting to use Big Data to obtain their customers insight in order to serve them in a better way. The use of Big Data has become quite a crucial way for businesses to compete with their competitors. Also not only companies gain from the value of Big Data, it is also the customer's hugely benefit from its usage. In association with Big Data's real time information, which is one of the most heavily used application of personal and location data. As there is a significant growth in the use of smart phones and the use of GPS services from the phones and other devices, the use of smart traffic routing will definitely grow and in turn it will hugely benefit the customers. Big Data is not a single packaged technology, it is in general a platform consists of usage of various components to achieve a common goal. There are plenty of components available in the market for the businesses to customise their Big Data platform. The utilization of Big Data is becoming more and more essential to businesses and it is even more important for them to adopt the right Big Data platform to accomplish their goals. The main aim of this study is to propose a framework for building a Big Data platform for publishing industry. The proposed framework was validated in an UK based news publishing organisation to find out the suitability and adoptability of the framework for their Big Data platform.

Keywords: Big Data framework · News publishing · Big data platform · NoSQL databases

1 Introduction

According to the IBM research, we create about 2.5 quintillion bytes of data every day. This massive volume of data comes from variety of sources like social media sites and applications, smart phone usage activities, trading and transaction activities, digital pictures and videos etc.

Reference [8] this volume of data is very staggering but amongst this volume of data there lies an immense opportunity and potential to understand and foresee the future. The term Big Data is not new anymore to the organisations around the world. Lot of large organisations around the world have already started investing in their Big Data platforms [4]. The mid-size and small companies are starting to realise the potential opportunities from Big Data technologies. Although the organisations around the world understand the opportunities provided by the Big Data technologies, they still are very much mercurial on how to build their successful Big Data platforms. There are

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 377–388, 2015.
DOI: 10.1007/978-3-319-21009-4_29

plenty of vendors in the market place to provide Big Data platform solutions or to provide various Big Data components but understanding and choosing the right mix of components according to the organisation's Big Data platform strategy is very vital. Organisations without realising their core strategies and key components of their Big Data platforms will fail miserably whilst trying to build one. The main objective of this study is to identify and analyse the key layers and components required for a successful Big Data platform. For the purpose of this study we have developed a framework to create a Big Data platform for an UK based international news organisation.

2 Importance of Big Data and Analytics

The recent advancements in the Internet of Things, cloud computing, smart phones etc. is generating a vast volume of data while accessing a service. The organisations that provide services to their customers consume these generated vast volumes of data from varied sources and store them. Not all the generated data is persisted by an organisation but they do store the data that they think that would be valuable for them for their future analyses. When the volume of these generated data are in Megabyte or Gigabyte per day, the organisations would not have huge challenges in persisting or analysing these data. But, when the volume hits the range of Terabyte or Petabyte, it's a huge problem for any organisation to persist and analyse these data [3]. Big data is data that exceeds the processing capacity of conventional database systems. It doesn't fit the structure of the traditional database architecture. In order to gain insight value from this vast volume of varied data you must choose an alternative way of processing it.

Reference [15] according to IBM research, most of the companies in U.S. Have at least 100 TB of data stores, which is equivalent to 100,000 GB. As a result, in order to rapidly analyse these vast and varied data, organisations are in a need of a new technology. Big Data technologies are the perfect solutions for the above mentioned problem [16]. According to Bain & Company, one of the world's leading management consulting firm, after studying more than 400 large companies they have found that those who have adopted the Big Data technologies and advanced analytics capabilities are clearly outperforming their competitors by wide margin. The adopters are five times faster in making solid and good decisions when compared to their competitors, due to their insight knowledge obtained from the data analytics. It would help them to report a good financial performance with in their industries [17]. Big Data technologies and its related analytics have become increasingly paramount among the business communities and academics in recent decades. Many of the industrial research and studies have highlighted this major development.

In a survey conducted by Bloomberg Businessweek (2011), majority of the companies with revenues well over $100 million were to found to use some form of analytics to support their business decisions. The advancements in web technologies means the traditional information systems are now no longer used only for support systems for business. They have become a bare-bone for any digital businesses. The streams of user generated data in the forms of logs from these system gives insight not only about the systems operations but also about the behaviour patterns of the systems users.

Reference [13] the importance of Big Data and analytics can be very well understood with the experience of McLaren F1 team. When a McLaren F1 car is on the track, more than 120 sensors transmit a torrent of information on tire pressure, torque, temperature, and down force etc. to understand and analyse the state of the car. This shows how the company is data-obsessed in their racing operations to make a solid decision and to stay ahead of the competitions. McLaren launched a consulting firm based on their big data analysis and simulation expertise to help the other businesses outside the racing. They have helped to design a better health monitoring systems for sick children, created a scheduling system based on simulations for Heathrow Airport to reduce the flight delays. The simulation software created by McLaren for Heathrow Airport has improved the on-time airport performance by 19 %. McLaren is also working with the Britain's Olympic team to understand and analyse their performances or the athletes to further tune them. By doing so over the three Olympics the athletes worked with McLaren has won 17 gold medals, and 32 medals in all.

Another company who have pioneered the use of Big Data technologies and analytics is without doubt the Amazon Inc, the online retailer. Amazon is well known around the globe for their pioneered e-commerce business. There are other sub systems which help Amazon to improve their sales dramatically one of which is their personalised recommendation system, which is built on the big data that it acquires from its millions of customers activities. It is clearly becoming more evident not to ignore the importance of Big Data and the analytics in the current competitive business environment.

3 Identifying and Analysing the Key Components of a Big Data Platform Framework

According to the white paper 'Architecting a Big Data Platform for Analytics' published by IBM [1], it is clear that the Big Data is not a single packaged technology. Big Data is a platform and it consists of various layers of workflow and components. In order to understand these layers and components, we need to study the existing structure of the news publishing industry [5]. According to Financial Times Managing Director, Rob Grimshaw, "There are only a couple million people in the world who are genuinely relevant to our proposition" This led to the FT's belief that they have to focus on content-led business rather than volume-led. Rob Grimshaw also claims that they have produce quality and in-depth journalism content that can't be found anywhere else. If the people are willing to pay for the FT content then its journalism has to be too good. Of course their 60 % of FT.com's revenues comes from subscriptions. Even the FT's free content access areas of the site, such as the Alphaville blog, are perfectly fact checked and carefully edited. It is clearly evident that the valuable assert for any publishing industry is their content. It is their core assert based on which the whole business runs [12]. It's vital to understand the various content and data generated in publishing business and classify them into structured (well structured with relations), semi-structured (data with meta info) and unstructured categories.

Authors and Content

Author databases – The information about the authors (structured)
Content database – Rich content (structured)
PDF - documents (structured)

Metadata – Data about the data (structured)
Image libraries – Images used in the content (semi-structured)
Adverts – Advertising information (semi-structured)
Feeds – Content feeds (semi-structured)
Info Graphics – Data presentation (unstructured)
Audio – Audio files (unstructured)
Video – Video files (unstructured)
Content sharing – Content sharing on social media etc. (unstructured)
Ratings – Author, content and other resource ratings (unstructured)
Email – Communication documents (unstructured)

Data Providers

Affiliations – Affiliates information (unstructured)
Contracts – Data providers contracts (unstructured)
Payments – Third party payment information (unstructured)

Marketing and Sales

Geographical performance – Geographical sales performance (structured)
Campaign data – Marketing campaign information (unstructured)
Events – Marketing events (unstructured)
Surveys – Research surveys (unstructured)
Offers – Sales offers information (unstructured)

Readers/Consumers

CRM Data – Customer relationship management data (structured)
Subscriptions – Customers subscriptions information (structured)
Demographics – Customer demographics information (semi-structured)
Service Logs – Application service logs (semi-structured)
Customer preferences – Customer application preferences (structured)
Searches – Customer search actions (unstructured)
Content likes – Customer content likes (unstructured)
Content sharing – Customer content sharing information (unstructured)
Ratings – Customer content ratings (unstructured)
Content reading – Customer reading history (unstructured)

Almost 80 % of the available data are in unstructured category. It is important to clearly understand these data workflows and their historical importance for the businesses. There could be abundance of noisy data within the collected stream of data [19]. There are two common practices in dealing with heterogeneous data mining process. One is to enhance the data quality at the source or simply gather the noisy data and feed them to the process and filter them later during processing. There are various challenges to capture and organise the data from these heterogeneous systems. Some of the ways to capture the data flow from these existing systems is by employing the enterprise messaging service techniques or simply reading and parsing the log files of the system. Thus collected enormous data has to be stored in suitable database systems. There are wide varieties of database storage available for Big Data. Organisations that are willing to adopt the Big Data technologies are now collecting the huge volume of data are increasing adopting non-relational database, called NoSQL databases. NoSQL databases are largely for processing large scale datasets; it offers high horizontal scalability compared to the traditional relational database systems. Some of the popular NoSql databases are Google's BigTable, HBase, Cassandra, Voldemort, CouchDB, MongoDB, Redis, Riak, Neo4j etc. [11]. NoSQL databases are classified into four different categories.

Key-Value Stores

Key value stores are suitable for highly scalable retrieval of the values for an application. Amazon employs their own K-V system called Dynamo for some of its core services to provide highly available and scalable distributed data store. Other popular K-V systems are Voldermort, Redis, BerkeleyDB, Riak etc. [6, 7, 11, 14].

Document Databases

References [6, 14] Document Databases are designed to manage and store documents. The documents are encoded in a standard data exchange format such as JSON (Javascript Option Notation) or BSON (Binary JSON) or XML. Both the keys and values are fully searchable in document databases wheres in key value store this is not possible. Document databases are very useful for storing and managing text document, email messages, XML documents. They are also good for storing sparse data and de-normalised representation of database entity such as a customer or product. MongoDB and CouchDB are good examples of document database which are open source and schema free.

Wide-Column Stores

Wide-column stores similar to document databases apply a distributed, column oriented data structure that accommodates multiple attributes per key. Some Wide- column stores have a Key-value DNA which were patterned after Google's Big Table, developed mainly for Google search index, Google Earth and Google Finance. This type of DMS is mainly used for distributed data storage especially for WC/CF timestamping functions. They are also great for large scale data processing, sorting, parsing conversions between hexadecimal, binary and decimal code values. Big Table use by google, Hyper table, Cassandra used by Facebook, twitter and SimpleDB used by Amazon are some of the examples of Wide-Column/Column-Family stores [6, 14].

Graph Databases

Graph databases succeed relational tables with structured relational graphs of interrelated key-value pairings. They are similar to object-oriented databases as the graphs are described as an object-oriented network of nodes, properties and node relationships. Graph databases are quite useful when we are more concerned in relationships between data than in data itself. These are optimised for relationship traversing and not for querying. Some of the examples of Graph Databases are InfoGrid, Neo4j, InfiniteGraph, Sones GraphDB and AllegroGraph [6, 14].

Upon persisting the data we would need a highly efficient distributing data processing system to churn the data. We should be amazed with the fast response we get while searching the web with Yahoo or Google and how they manage to search millions of pages and gives back results in less than milliseconds. Both these giant search engines are driven by the algorithms originated with Google's Map Reduce Framework [9]. Map Reduce is a programming model which is an associated implementation for rendering big data sets that is improved to a vast variety of real world projects. Users define the calculation in terms of a map and a reduce function. The fundamental runtime system automatically parallels the computation across large scale clusters of machines. It also manages machine failures, and schedules communication between machines to make efficient use of the disks and networks. More than 10,000 distinct Map Reduce programs have been implemented across Google as programmers find it very easy to use. Apache foundation has implemented their own map-reduce open source framework called Hadoop, while Map Reduce is a proprietary technology. Apace has developed, Hadoop an open source platform for scalable, reliable and distributed computing. Hadoop is highly scalable from a single server to thousands of machines, each offering local storage and computation. Hadoop software library is a framework which allows distributed processing of big data using simple programming models across clusters of computers. It is designed to sight and manage failures at an application level, in order to deliver a highly- available service on top of a cluster of computers, each of which may be inclined to failures. Yahoo and many others are using Hadoop which successfully processes large amounts of data easily. These data processing units can be implemented as a micro service for specifically fitting a specific business area and in turn all the micro services could be integrated using the aggregated integrated service to provide various web services to its downstream systems.

Any distributed processing and data storage requires a dynamic infrastructure environment to facilitate easily creation of the required processing nodes for its processing needs. The cloud service infrastructure is the perfect match for these requirements. There are many popular cloud service providers in the market and Amazon web services is notably the pioneer in providing this service. Amazon web services provide a mixture of cloud- based computing. It also includes a wide selection of compute instances that can scale automatically (up and down) to meet the needs of the applications, with load balancing service and fully managed desktops in the cloud [2].

The processed data has to provide clear and concise value to the business and hence the reporting the analytics data in terms of suitable visualisation becomes a paramount importance. Visualising data is very important and common in today's world [10]. Data visualisation is mainly used to create and visually represent data points to provide enough information to the investors, progress reports and even visualise ideas models for customer segments. Data visualisations are a valuable tool available in a variety of features. Most of these tools are an Open - Source application that can be used in combination with one another or using JSON, Java Script, SVG, Python, HTML5, or drag and drop functionality, with the existing design application. Other tools available in the market are a wide-ranging business intelligence platforms with an ability to understand sophisticated analysis of data and reporting of data. D3js, Data driven Documents is a Java script library for changing documents based on data D3 assists you bring the data to life using HTML, CSS, and SVG with full capabilities of modern browsers and restrictions of proprietary framework. Key features of D3js are creating interactive SVG bar charts and binding uninformed data to DOM. It also, helps generate HTML tables from data sets. D3js has a variety of components and plugins to enhance facilities with built-in reusable components for ease coding [18]. For example the Atlas project of MIT institute has used D3js library to build the Observatory of Economic complexity visually. The Observatory of Economic Complexity makes international trade data and economic complexity indicators available through millions of interactive visualizations (http://atlas.media.mit.edu/). The other popular library which fits for the same purpose is Lumify, which is an open source Big data analysis and visualisation platform. It helps us analyse relationships with data, automatically discover paths between entities, and set up new links in 2D or 3D. It helps organise work into separate workspaces that can be shared with colleagues. It also, helps us overlay data as layers on a map for a geographical view of data model.

Therefore based on the literature review we have the proposed layers and components of the Big Data platform framework illustrated in the below Figs. 1 and 2:

As part of this study we have tried to identify and analyse the key layers and components of a Big Data platform specifically for a publishing industry. We have developed the following hypothesis scenario based on this proposed framework and have validated it with an UK based international news organisation.

Hypothesis 1: The proposed Big Data platform framework could be adopted easily across the wider section of the business to deliver business value.

Hypothesis 2: All the layers and components are clear and complete to build a Big Data platform.

Hypothesis 3: The proposed framework is very cost effective for adaptation.

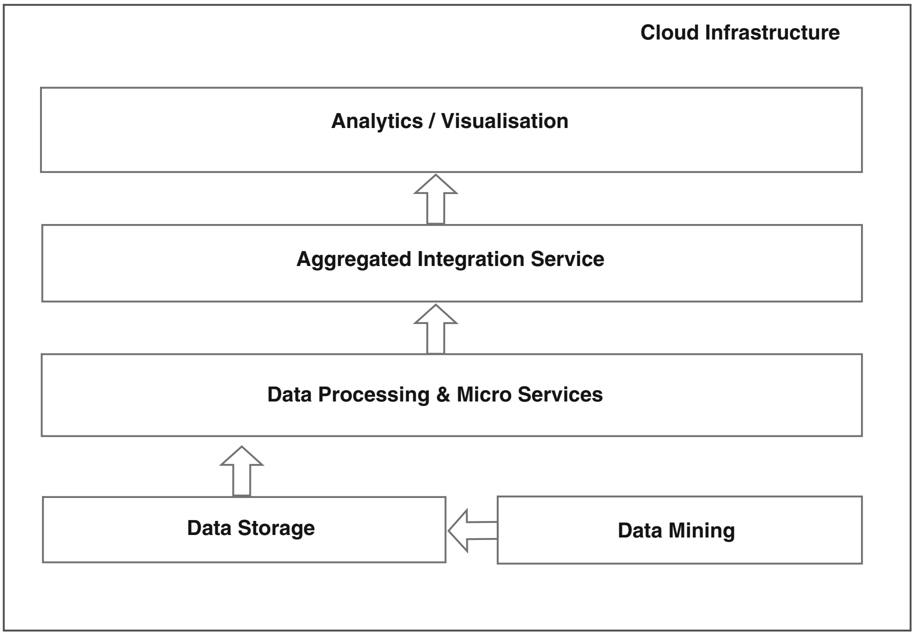


Fig. 1. Big Data platform layers

Layers	Components
Data Mining	New Data points - Enterprise messaging service, log file processors. Existing Data - Migration
Data Storage	Key Value NoSQL database, Document databases, Wide-Column stores and Graph Databases.
Data Processing & Micro Services	The Apache Hadoop MapReduce (Distributed data processing), Micro Services (Real time data providers)
Aggregated Integration Service	Aggregated Integration service to aggregate the micro services.
Analytics / Visualisation	Business Query processors, Visualisation tools
Cloud Infrastructure	Cloud Service infrastructure (easily create and deploy environments).

Fig. 2. Big Data platform components

4 Research Method

This study uses the sample survey responses from employees of an UK based news publishing company. A total of 54 employees volunteered in this study. The participants agreed to adopt the proposed framework by trying to implement a proof of concept system for their new recommendation engine and feedback their views in the provided survey. An one hour dedicated presentation session was provided to all the participants to explain the proposed framework. We have also agreed to a timeframe of 4 weeks to implement the proof of concept. Personal level demographics information shows 60 % participants holds management positions with in the company and the remaining 40 % are software engineers. Data was collected using a survey questionnaire with about 30 measuring items. The questionnaire was divided into three parts: The first part measuring the framework usability/adoptability with three dimensions including big data usability intention, framework clarity, framework suitability and 10 items. The second part measuring the framework completeness with two dimensions including layers/components maturity, framework complexity and 10 items. The final third part measuring the framework cost effectiveness for adoption with three dimensions including the skills requirements, tools, time requirements and 10 items. In addition, there were also four variables pertaining to the participant's personal data including position, gender, education and seniority.

4.1 Reliability and Validity Analysis

An explanatory factor analysis was carried out to ensure the validity of all the measures. The Table 1 above illustrates the variable factor loadings of the items belonging to the three key dimensions framework usability/adoptability, framework completeness and framework cost effectiveness respectively.

5 Hypothesis Testing Using Descriptive Analysis

We have adopted the descriptive analysis method to test the stated three hypotheses. Hypothesis 1 argues for the easy adoptability of the proposed Big Data platform framework across the wider section of the business to deliver the business value. Although 70 % of participants agree to the suitability of the proposed framework, which certainly supports the hypothesis 1. But almost 50 % of them lack certain clarity in the need of few proposed components. Especially the usage of micro services and aggregated integrated services were not clear for the adopters. Also 85 % of respondents don't see any immediate benefits in Big Data technology usage with in their programme

Table 1. Exploratory factor analysis results

Dimensions	Items	Factor loadings	Dimensions	Items	Factor loadings
Framework usability/adoptability			**Framework cost effectiveness**		
Big data usability intention	100 101 102 103	0.873 0.554 0.430 0.685	Skills requirements	120 121 122	0.467 0.452 0.340
Framework clarity	104 105 106	0.785 0.549 0.775	Tools requirements	124 125 126	0.744 0.352 0.420
Framework suitability	107 108 109	0.890 0.777 0.750	Time requirements	127 128 129 130	0.565 0.650 0.363 0.435
Framework completeness					
Layers/components maturity	110 111 112 113 114	0.455 0.558 0.675 0.780 0.560			
Framework complexity	115 116 117 118 119	0.445 0.760 0.850 0.772 0.665			

stream. Hypothesis 2 argues that all the layers and components are clear and complete to build a Big Data platform. Most of the 65 % of respondents have not agreed with the completeness of the framework and they felt its more generic than specific for their use case. Although, 85 % respondents agree with the proposed layers and components structure. In general they felt the generic components proposal is very complex and time consuming to further evaluate the specific tools. Even to choose a NoSQL database from the available four categories they required to evaluate various products to test their suitability. Also, lot of software engineers were not happy in using log processors as part of data mining as they feel it has a high dependency on the logging patterns in the application. These arguments clearly not supporting the hypothesis 2 and which illustrates the proposal should include specific technology templates as part of the framework for specific sectors. Hypothesis 3 argues for the cost effectiveness of the proposed framework. Almost 72 % of respondents complained about the skills required implementing the framework, they felt it's more complex and require training to improve the skills or they need new resources with this new set of skills to perform the required actions. Again 70 % of respondents reported the complexity in using these new tools takes lot of time and requires a steep learning curve. 75 % of respondents felt the agreed 4 weeks' time is not adequate to complete the proof of concept. These arguments were against the hypothesis 3 and it illustrates the complexity involved in implementing these new Big Data technologies in an organisation. Also, it clearly shows the requirement of the new breed of skills to implement these technologies.

6 Limitations

This study has many limitations. The proposed Big Data platform is very generic and does not suggest any specific set of tools. We have tried to adopt this proposed Big Data platform as a proof of concept in a UK based news publishing company and also there is a limited time constraint to implement this framework properly. This study did not a have a separate data scientist allocated for the proof of concept and thus the analytics were not generated adequately. Data scientist plays a key role in generating the required information and interpreting the obtained information in a concise way to the key stakeholders. We have collected the data from only one news publishing organisation, the results cannot be generalised for all industry sectors due to a smaller sample size. The participants did not have enough time to understand and analyse the various available technology components in various layers. Overall, a data driven approach is limited not only by quantity and quality of the available data, but also by the chosen boundary specifications, metrics and analysis methods as well as visual presentation methods.

7 Recommendations

As this study is the first step in defining the framework for the Big Data platform it needs further work in various layers and components. It would be more appropriate to undertake further studies to analyse the available tools in various layers and suggest a best combinations for developing a Big Data platform for organisations. Almost 85 % of respondents in this study don't see any immediate benefits in implementing the Big Data technologies, which is quite shattering. Although they are much aware of the Big Data technologies but they are not willing to focus on the long term benefits instead prefer to achieve the short time goals. The benefits of the Big Data technologies should be promoted with in the organisations. The adequate opportunities should be provided for the employees to evaluate these technologies where ever required. The technology components would change based on the programme stream requirements, so research time should be allocated to find the relevant technology to address their specific business problems. One of the best methods to adopt any new technology is to try with the new systems. It is a good opportunity to prototype these new systems with the Big Data technologies and test using the existing infrastructure and available open source free solutions in the market. It is also very important to balance the existing technology infrastructure with the introduction of new Big Data technologies. It needs a lot of efforts and careful consideration while migrating the old existing systems to a newer one.

8 Conclusions

To succeed in today's data driven business environment requires being able to think about how these fundamental concepts apply to their particular business problems. There is a strong evidence that business sustainability and performance can be improved via data-driven decision making. Depending upon the Big Data framework the decision support system sometimes also allows making decisions automatically at a massive scale.

In the era of charging for the content by the publishers, to stay ahead in competitions, they need insight knowledge into their data analytics. It is becoming very evident that the publishers cannot afford to overlook the big data and analytics. Also, a successful implementation of Big Data strategy would lead an organisation to innovate in various fronts. It is very evident that Big Data shouldn't be considered as a single packaged technology rather it needs to be thought as a platform comprising of various layers and components. The result of the study iterates the suitability of the proposed framework for building a Big Data platform with lot of other potential opportunity for further improvements in the overall framework. As suggested by the participants, it would be important to suggest the technology templates for specific use case needs along the framework would be a good enabler for adapting this framework. This framework if expanded could be perhaps used in different industries as well. The other important outcome of this study is the lack of understanding of the employees about their companies Big Data strategies, which clearly impedes the employees contributions. Although the employees motivation towards contribution for Big Data platform is not in scope of the current study but it would be essential to undertake such studies in this area. There is no doubt that we need further studies in this area to progress and amend the layers and components in the proposed framework.

References

1. Architecting a Big Data Platform for Analytics: Architecting a Big Data Platform for Analytics (2015). http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=SA&subtype=WH&htmlfid=IML14333USEN. Accessed 13 April 2015
2. AWS|Amazon Elastic Compute Cloud (EC2) - Scalable Cloud Hosting: AWS|Amazon Elastic Compute Cloud (EC2) - Scalable Cloud Hosting (2015). http://aws.amazon.com/ec2/. Accessed 20 February 2015
3. Big data: the next frontier for innovation, competition, and productivity|McKinsey & Company: Big data: The next frontier for innovation, competition, and productivity (2015). http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation. Accessed 18 January 2015
4. Dumbill, E.: Planning for Big Data: A CIO's Handbook to the Changing Data Landscape. O'Reilly Media, Sebastopol (2012)
5. Eight Lessons from the Financial Times' Digital Success|Mediashift|PBS: 8 Lessons from the Financial Times' Digital Success|Mediashift|PBS (2015). http://www.pbs.org/mediashift/2014/05/8-lessons-from-the-financial-times-digital-success/. Accessed 10 February 2015
6. DeCandia, G., et al.: Dynamo: amazon's highly available key-value store. In: SOSP 2007, Proceedings of Twenty-First ACM SIGOPS, New York, USA, pp. 205–220 (2007)
7. Han, J., Haihong, E., Le, G., Du, J.: Survey on NoSQL database. In: 2011 6th International Conference on Pervasive Computing and Applications (ICPCA), pp. 363–366. IEEE, October 2011
8. IBM big data platform - Bringing big data to the Enterprise: IBM big data platform - Bringing big data to the Enterprise (2015). http://www-01.ibm.com/software/data/bigdata/. Accessed 16 January 2015
9. Dean, J., Ghemawat, S.: MapReduce: simplified data processing on large clusters. Commun. ACM **51**(1), 107–113 (2008)
10. Steele, J., lliinsky, N.: Beautiful Visualization: Looking at Data through Eyes of Experts. O'Reilly Media Inc., Sebastopol (2010)
11. Konstantinou, I., Angelou, E., Boumpouka, C., Tsoumakos, D., Koziris, N.: On the elasticity of NoSQL databases over cloud management platforms. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, pp. 2385–2388. ACM, October 2011
12. Tewari, K.: Leveraging Big Data Opportunities For Growth – Keynote (2013). Online Video, 22 April. http://www.frequency.com/video/krishna-tewari-leveraging-big-data/92326953/-/5-1519156. Accessed 21 February 2015
13. McLaren Uses Racing Expertise in Data-Driven Consulting – Businessweek: McLaren Uses Racing Expertise in Data-Driven Consulting – Businessweek (2015). http://www.bloomberg.com/bw/articles/2014-10-02/mclaren-uses-racing-expertise-in-data-driven-consulting. Accessed 10 February 2015
14. Padhy, R.P., Patra, M.R., Satapathy, S.C.: RDBMS to NoSQL: reviewing some next-generation non-relational databases. Int. J.Adv. Eng. Sci. Technol. **11**(1), 15–30 (2011)
15. Pethuru, R.: Handbook of research on cloud infrastructures for big data analytics (2014). Information Science Reference (an inprint of IGI Global)
16. Thomas, R.: Big Data Revolution: What Farmers, Doctors and Insurance Agents Teach Us About Discovering Big Data Patterns, 1st edn. Wiley, New York (2015)
17. Russom, P.: Big data analytics. TDWI best practices report, 4th Quarter 2011 (2011)
18. Ghemawat, S., Gobioff, H., Leung, S.T.: The Google file system. In: ACM SIGOPS Operating Systems Review, vol. 37, pp. 29–43. ACM (2003)
19. Vilas, K.S.: Big data mining. Int. J. Comput. Sci. Manage. Res. **1**(1), 12–17 (2012)

Cloud Based Office Suites - Users' Attitudes Towards Productivity, Usability and Security

Reinhard Bernsteiner(✉) and Robert Pecina

Management Center Innsbruck, Innsbruck, Austria
reinhard.bernsteiner@mci.edu, robert_pecin@yahoo.de

Abstract. Cloud Computing is one of the most important developments in the field of information technology. This new kind of service delivery also offers innovative opportunities for office suites usually called Office 2.0. Those tools allow users to access their data and documents anytime and anywhere, as long as an internet connection is available. Due to the internet-based design these tools can be used on almost any device which has a browser installed.

In the first section, the theoretical basis for cloud-based office suites is discussed. The second part presents the results of an empirical survey conducted by the authors and explores the attitudes and feelings towards usability and security of those tools. An experimental setting was used to gather data from the participants. Google Docs was selected to conduct the experiment.

Keywords: Office 2.0 · Google Docs · Usability · Security · Experiment · Empirical study

1 Introduction

Nowadays many users, companies as well as individuals, need to have access to their data and documents anywhere and anytime on all types of devices. The same is true for the tools and applications that are necessary to work with those data and documents. Cloud based services can be used to fulfil these requirements.

The concept of cloud computing is one the emerging technologies which will massively change the way how IT services will be provided in the future. One of these cloud based services is office suites. Services, data or documents can be accessed as long as an internet connection is available, meaning that these elements reside somewhere on serves that cannot be touched physically. These servers are in datacentres which are not on premise and thus are out of personal control.

This means that applications and data have to be transferred over the internet from the cloud to the personal device or processed by the cloud based services. This represents the foundation of Office 2.0 and as a consequence leads to concerns about data security, productivity and usability. For this investigation Google Docs was selected as a commonly used Office 2.0 tool.

Cloud based services, especially Software as a Service Services (SaaS) gain more and more in importance, in private as well as in professional matters. The number of companies using Office 2.0 software as their key asset for collaboration or editing data or documents is growing. This trend justifies the relevance of this research endeavour.

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 389–403, 2015.
DOI: 10.1007/978-3-319-21009-4_30

The central aim of this work is to find out more about the users' experience concerning data security, productivity and usability of Office 2.0 tools. Google Docs, which is widely used, was selected as a representative of an Office 2.0 tool.

2 Theoretical Background

The following chapters refer to the necessary theoretical background required for the following empirical survey.

2.1 Office 2.0 and Web 2.0

The term Web 2.0 was introduced by Tim O'Reilly, who suggested the following definition:

"Web 2.0 is the network as platform, spanning all connected devices; Web 2.0 applications are those that make the most of the intrinsic advantages of that platform: delivering software as a continually updated service that gets better the more people use it, consuming and remixing data from multiple sources, including individual users, while providing their own data and services in a form that allows remixing by others, creating network effects through an "architecture of participation", and going beyond the page metaphor of Web 1.0 to deliver rich user experiences. The focus of Web 2.0 is on the behaviour of the user. It should empower people to communicate, collaborate, contribute, and participate" [1].

In today's economy and for people in offices who drive it, collaboration, contribution and participation are key elements in their daily routines. Major parts of today's office work are supported by stand-alone office applications. Office 2.0 software addresses the issue of bridging office applications between users so that they can collaborate in a better way thereby improving personal and organizational productivity. Today's Office 2.0 applications do not simply focus on collaborative data editing but also on data sharing, its review and on document resource management. Office 2.0 software usually provides all these abilities without forcing the users to set up their own IT infrastructure. Thus, the ideas of Office 2.0 show many parallels to the principles of Web 2.0.

Due to globalization and increasing competition, the needs for more flexibility are requirements for office tools that gain in importance. In addition to flexibility mobility is another factor that influences the future developments of Office 2.0 usage. Today users of office applications want to be able to have access to their data with their personal devices and their personal IT infrastructure at any time and from everywhere. Office 2.0 applications make it possible to cope with these requirements by storing and providing applications, data and documents in the cloud [2].

By using different operating systems on different devices as well as the variety of available office tools problems regarding the compatibility or transferability of documents arise. Thus errors are much more likely to occur. Those problems can be reduced by using a generally available office solution that is independent from operating systems and which is available on all kind of devices.

Platforms on which users can share information, their opinions, or revise already existing documents or other artefacts are needed [4, 5]. This leads to the ideas which go in line with the concepts of Web 2.0, joint effort, collaboration and cooperation.

For example Microsoft OneDrive or Google+ represents examples of Web 2.0 technologies. This indicates that Web 2.0 is the second generation of the Web [3]. *"Web 2.0 technologies have caused three effects:*

1. *A shift in locus of activity from the desktop to the Web*
2. *A shift in locus of value production from the firm to the consumer [...]*
3. *A shift in the locus of power away from the firm to the consumer.*

Indeed, Web 2.0 can be thought of as a series of technological innovations in terms of both hardware and software that facilitate inexpensive content creation, interaction, and interoperability, and that put the lay user - rather than the firm - center stage in terms of design, collaboration, and community on the World Wide Web" [6].

2.2 Cloud Computing

According to the NIST definition, *"cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction"* [7].

From a user's point of view there are three main delivery models of cloud computing. The most commonly used delivery model is Software as a Service (SaaS). In this model applications are provide in the cloud, eliminating the need to install and run the application on premise. A Platform as a Service (PaaS) facilitates the development and deployment of applications. Infrastructure as a Service (IaaS) is the most basic cloud service model in which computers, physical or more often virtual machines and other resources are offered. This method of outsourcing computing capacities brings advantages to users of devices with low computing resources like netbooks, laptops or smart devices like smart phones or tablets. Using cloud based services the devices only have to establish and keep up a network connection and do not have to process the software on the device. This implies the disadvantage that the software and relevant data or documents are only available as long as the computer has access to the internet [8].

Cloud computing services are usually rented [8]. Users only pay for the actual usage of the services (pay-per-use). This enables companies to turn capital expenses to operational expenses, because upfront investments in IT infrastructure or software can be reduced [9, 10]. Cloud based services are more easily scalable compared to stand-alone software because the resources in the cloud are usually shared by several tenants [9, 11, 12].

In the earlier days of cloud computing there were restrictions in terms of the bandwidth of the required internet connection. Nowadays the available networking infrastructure is usually powerful enough to cope with these demands.

Another concern in the context of cloud computing is security risks. These risks have their origin in the transfer of data and commands from the used device to the cloud as well as in the execution of the commands or processing of data in the cloud [13, 14].

2.3 Google Docs Representing Office 2.0

It is the intention of Google to offer with Google Docs a cloud based office suite as an alternative to the wide spread stand-alone office applications. Software as a Service (SaaS) is the cloud service model for Google Docs.

Google Drive is the platform from which all of Google's applications are started. Google Drive offers the opportunity to store and edit data or documents in the cloud as well as to share data with other users. There are several options to set access rights to create, change or delete documents. Another feature of Google Docs is that several users can edit the same document at the same time and these changes are visible in real time.

In terms of availability, Google guarantees 99.9 % uptime and due to security concerns encrypts all data transferred to or from Google Drive as well as data being stored on it [15].

The terms of use concerning of data confidentiality need to be taken into consideration:

"Some of our Services allow you to upload, submit, store, send or receive content. You retain ownership of any intellectual property rights that you hold in that content. In short, what belongs to you stays yours.

When you upload, submit, store, send or receive content to or through our Services, you give Google (and those we work with) a worldwide license to use, [...] such content. [....] This license continues even if you stop using our Services" [16]. Thereby the principle of data confidentiality is not provided. All uploaded data or documents can be used by Google without notice. It is not advised to upload sensible data which are worth protecting because intellectual property rights are not guaranteed.

Google Docs offers applications to edit files containing text, tables, presentations, forms, graphics or maps. Documents with common data formats can be opened and viewed by the service as well.

3 Empirical Survey

In the present survey an experimental design was used to explore the attitudes with respect to the topics productivity, usability and the feeling of data security of Office 2.0 suites. Google Docs was selected as an Office 2.0 representative for this survey.

The experiment was partly conducted online in order to provide a more flexible interaction with subjects as well as to standardize the data collection process.

According to Crano & Brewer, experimental research involves *"systematically controlling the variation in the independent variable (or variables) to assess its causal impact"* [17].

3.1 Experimental Design and Target Group

First, the participants were introduced to the study and the aim of the study. In order to have a predefined system environment, accounts for Google Docs were provided to the participants. Each participant received a personal account allowing several participants to conduct the tasks at the same time. Then the participant had to fulfil the tasks with the tools. After having finished the tasks, the participants had to fill out a questionnaire regarding the topics productivity, usability and the perception of data security.

The questions of the questionnaire were taken from two certified surveys developed by the University of Osnabruck, faculty of psychology for the evaluation of graphical user interfaces [18, 19].

The questionnaire consists of three parts. The first two parts comprise questions regarding the perception of security, working with the software, suitability for daily use and usability. The third part consists of demographic questions. There was no time limit for the participants to fill out the questionnaire.

The target group of the experiment are Central European students that already have experiences in the IT-sector or have some education in the IT-Sector. In terms of age the participants showed a wide range in their age, mostly between 19 and 24. Details are depicted in the following (Table 1):

This target group was specifically chosen because today's students are the workforce of tomorrow. It can be assumed that students of today are adopters of Office 2.0 tools once they enter the labour market. Since some of them are already using Office 2.0 software today, they will influence further developments and the usage of this kind of software in the future.

After the first design of the experimental procedure a pre-test was done. Some changes on basis of the feedback were integrated in the final design of the experiment.

At the end of the investigation phase all data gathered were imported into SPSS and then statistically analysed. All unfinished data entries have been deleted in order to provide data integrity. The survey was conducted from December 2014 to mid-January 2015. The results are presented consequently in tables only.

Table 1. Age of participants

		Frequency	Percentage
Valid	18 to 21	58	56,9
	22 to 25	33	32,4
	over 26	11	10,8
	Total	102	100,0

All in all 102 individuals participated in the experiment and finished all parts of the experiment, including the questionnaire.

3.2 Tasks for the Participants

The participants had to fulfil three major tasks in order to explore the software and get valid results:

1. The participants had to correct ten spelling-mistakes that have been integrated in a given text that has been stored on the individual software-environment.
2. They were also instructed to change the layout of the text by applying different fonts, colour the headline of the given text, etc.
3. Furthermore they were ordered to reproduce a graph which was provided in form of a picture.

Two different documents were provided for the tasks of the experiments in order keep learning effects low. Both provided texts were written in English. English was used to be able to repeat the experiment on an international level. Both texts had a length of around 190 words, again in order to keep the learning curve low.

The participants had a timeframe of ten minutes to complete the whole experiment. After ten minutes they were asked to stop their work and fill out the questionnaire. Participants who were able to finish before the ten minute threshold were asked fill out the questionnaire immediately after having finished the experiment.

The questions had to be answered on a 5-point Likert scale. The participants hat not the option to skip questions or to state that they do not want to answer specific questions. The response options did not include answers like “I don’t know” or “Does not apply”.

3.3 Hypotheses

In this section, the hypotheses for the research are presented. The respective results are shown and discussed.

Hypothesis 1. The percentage of students that feel as if the software forces them to perform tasks that are not related to their actual work is beneath 20 % (Table 2).

Table 2. Evaluation of non-work related tasks

The software forces me to perform tasks that are not related to my actual work.				
		Quantity	Percentage	Cumulative percentage
Valid	absolutely agree	5	4,9	4.9
	generally agree	16	15,7	20.6
	neither…nor	36	35,3	55.9
	slightly disagree	34	33,3	89.2
	absolutely disagree	11	10,8	100.0
	Total	102	100,0	

20.6 % of the participants either partly or wholly agreed with the statement. When looking at the table from the flip side, it states that 44.1 % had none or few problems handling the software and 35.3 %have a neutral position. It can be said that this hypothesis can be rejected because there are 44.1 % who do are not forced to perform additional tasks.

This could be traced back to the tasks that the participants had to do during the experiment. In order to determine this possibility question two was integrated in the questionnaire. An analysis of correlation upon Q1 (The software forces me to perform tasks that are not related to my actual work.) and Q2 (How easy was it for you to perform the given tasks?) should shine light on the problem which is presented in the next table (Table 3).

Table 3. Correlation of given task with forced additional tasks by software

Analysis of correlation			
		Q1	Q2
Q1	Pearson correlation	1	-.087
	Unilateral significance		.193
	N	102	102
Q2	Pearson correlation	-.087	1
	Unilateral significance	.193	
	N	102	102

The analysis of correlation shows that there is no significant correlation between the answers of the participants of the two questions. This means that the tasks that had to be fulfilled and the usability of the software are independent from each other.

Another explanation could be the high complexity of the software-handling. Analysing the correlation between the two questions Q3 (Do you think that Online Office products are more complicated to work with than Stand Alone products?) and Q1 (The software forces me to perform tasks that are not related to my actual work.) can shed light on this aspect (Table 4):

Due to the significance of 0.259 this hypothesis cannot be rejected. But the analysis shows a very weak correlation. It can be said that the complexity of handling the

Table 4. Correlation forces additional tasks and handling-complexity

Analysis of correlation			
		Q1	Q3
Q1	Pearson correlation	1	.065
	Unilateral significance		.259
	N	102	102
Q3	Pearson correlation	.065	1
	Unilateral significance	.259	
	N	102	102

Table 5. Time savings by avoiding to merge documents

Do you think that you could save time if merging documents created by different authors could be avoided?				
		Quantity	Percentage	Cumulative percentage
Valid	Absolutely agree	30	29.4	29.4
	generally agree	37	36.3	65.7
	neither…nor	23	22.5	88.2
	slightly disagree	12	11.8	100.0
	Total	102	100.0	

software and the difficulty of the tasks that had to be fulfilled are not influencing one another.

Hypothesis 2. More than 95 % think that time can be saved by avoiding the task of merging documents created by different authors Table 5.

Almost two thirds (65.7 %) of the participants answered either partly or wholly positively. Only 11.8 % think that joint editing of documents does not save time; 22.5 % have a neutral opinion. It must be said that the goal of 95 % was not reached but a majority of almost two thirds agree on this issue.

An explanation for this could be found in the comparison of the start-up of stand-alone office solutions and Office 2.0 software Table 6.

When reviewing the rating options in the 5-point Likert scale compared to the mean of the statistical analysis, one can see that 2.4 is not brilliant. Thus, the start-up of the software has to be judged complicated and time-consuming. The time savings that origin from the avoidance of merging documents which had to be written in the same file could thereby be compensated by the complexity and duration of the start-up.

This statement a can be validated with the analysis of correlation of the following two questions Q1 (Do you think that you could save time if merging documents created by different authors could be avoided?) and Q2 (Rate the start-up of the product in

Table 6. Start-up of Office 2.0 software compared to stand-alone solutions

Rate the start-up of the product in reference to the start-up of a stand-alone office product e.g.: Microsoft Office is started by a double click on an icon.				
		Quantity	Percentage	Cumulative percentage
Valid	good impression/easy	16	15.7	15.7
	slightly good impression	43	42.2	57.8
	neither…nor	30	29.4	87.3
	slightly bad impression	12	11.8	99.0
	bad impression/complicated	1	1.0	100.0
	Total	102	100.0	
Mean	2.40			

Table 7. Correlation of time savings and start-up

Correlation analysis			
		Q1	Q2
Q1	Pearson correlation	1	.165
	Unilateral significance		.049
	N	102	102
Q2	Pearson correlation	.165	1
	Unilateral significance	.049	
	N	102	102

reference to the start-up of a standalone Office product. E.g.: Microsoft Office is started by a double click on an icon.) Table 7:

The results show a weak positive correlation (0.165) between the two aspects with a significance of 0.049.

Hypothesis 3. More than 95 % of the participants are not happy with editing documents online and storing them in the cloud afterwards Table 8.

51 % of all participants were fine with the concept of editing documents online and storing them on a cloud based service afterwards. Adding the 20.6 % of participants that did not care about this functionality, this leads to a cumulative percentage of 71.6. 28.4 % were partly or completely concerned with editing and storing data online. It can be said that a majority is fine or neutral with online editing and storing documents in the cloud whereas the goal of 95 % was not reached.

The analysis of the answers to the question "Are you afraid of third parties reading your data without your knowledge" shines more light on the topic of attitude of the user towards data security Table 9:

The analysis shows that 64.7 % of all participants have doubts about data confidentiality whereas 22.5 % trust Google Docs to securely store their data without using these data or give access to third parties. This conflicts with the terms of use concerning data confidentiality published by Google.

Table 8. Attitude towards data security

How happy are you with editing documents online and storing them in the cloud afterwards?				
		Quantity	Percentage	Cumulative percentage
Valid	happy	22	21.6	21.6
	slightly happy	30	29.4	51.0
	neither…nor	21	20.6	71.6
	slightly concerned	15	14.7	86.3
	concerned	14	13.7	100.0
	Total	102	100.0	

Table 9. Attitude towards data confidentiality

Are you afraid of third parties reading your data without your knowledge?				
		Quantity	Percentage	Cumulative percentage
Valid	not afraid	12	11.8	11.8
	near to not afraid	11	10.8	22.5
	neither…nor	13	12.7	35.3
	slightly afraid	44	43.1	78.4
	afraid	22	21.6	100.0
	total	102	100.0	

Table 10. Acceptance of storage of sensible data that is worth protecting

Would you store sensitive data on this service?				
		Quantity	Percentage	Cumulative percentage
Valid	I would	4	3.9	3.9
	often, yes	10	9.8	13.7
	neither…nor	6	5.9	19.6
	barely	21	20.6	40.2
	I would not	61	59.8	100.0
	Total	102	100.0	

The next question refers to the issue if user would like to store data that is worth protecting on Google Docs. The intention of this question is to highlight the usage of Google Docs in the context of confidential or sensitive data Table 10.

Only 13.7 % would store sensible data that is worth protecting on the Office 2.0 solution and thereby partly or wholly edit and store them in cloud based services. The awareness of security leaks and risks with Office 2.0 applications seems to be existent. These concerns are ignored when it comes to data that is not worth protecting even if there is a possibility of tracing it back to the user.

Hypothesis 4. More than 80 % of the participants experience Google Docs to be user-friendly Table 11.

When reviewing the rating options on a 5-point Likert scale compared to the mean of the statistical analysis of 2.41 it can be said that the tool is rated slightly user-friendly.

60.8 % of the participants consider the application to be user-friendly. 26.5 % have a neutral perception. Only 12.7 % of the participants consider the tool as not user-friendly. All in all the answers of this question present a quite non-uniform perception of the user-friendliness. However, the hypothesis has to be turned down as 60.8 % are not enough compared to the required 80 %.

Table 11. Experienced usability

In your opinion, how user-friendly do you think is the product?				
		Quantity	Percentage	Cumulative percentage
Valid	very user-friendly	14	13.7	13.7
	Slightly user-friendly	48	47.1	60.8
	neither…nor	27	26.5	87.3
	slightly disagree	10	9.8	97.1
	not user-friendly at all	3	2.9	100.0
	Total	102	100.0	
Mean	2.41			

Table 12. Comparison of handling-complexity

Do you think that Online-Office products are more complicated to handle than stand-alone products?				
		Quantity	Percentage	Cumulative percentage
valid	a lot easier	6	5.9	5.9
	slightly easier	19	18.6	24.5
	neither…nor	47	46.1	70.6
	slightly more complicated	26	25.5	96.1
	way more complicated	4	3.9	100.0
	Total	102	100.0	

In order to get more insights regarding usability stand-alone software solutions can be compared with Google Docs Table 12.

46.1 % experienced the Office 2.0 software as equally complicated. Together with the participants that answered that the software is easier to handle, the percentage rises to 70.6 %. All in all, this means that Office 2.0 software is usable but still has room for improvements in comparison with stand-alone software solutions.

Hypothesis 5. The questions Q1 (How often do you think data is being edited and stored on this service and afterwards being read by third parties without the editor's knowledge or consent?) and Q2 (Are you afraid of third parties reading your data without your knowledge?) show significant correlation Table 13.

The analysis of correlation shows a weak negative correlation (-0,166) with a significance of 0.048.

The topic of data security can be explored further by an extended analysis of correlation: Q3 (How happy are you with editing documents online and storing them in the cloud afterwards?), Q4 (Do you fear that the service might lose your data?), Q5 (Would you store sensitive data on this service?), Q6 (How trustworthy / professional do you rate this product?) Table 14.

Table 13. Correlation of quantity and fear of data non-confidentiality

Analysis of correlation			
		Q1	Q2
Q1	Pearson correlation	1	-,166
	Unilateral significance		,048
	N	102	102
Q2	Pearson correlation	-,166	1
	Unilateral significance	,048	
	N	102	102

Table 14. Correlations of data security

Analysis of correlation							
		Q3	Q2	Q4	Q5	Q6	Q1
Q3	Pearson correlation	1	.327	.268	.292	.453	.083
	Bilateral significance		.001	.006	.003	.000	.407
Q5	Pearson correlation	.292	.333	.212	1	.188	-.003
	Bilateral significance	.003	.001	.032		.059	.979
Q6	Pearson correlation	.453	.248	.335	.188	1	-.109
	Bilateral significance	.000	.012	.001	.059		.275
	N	102	102	102	102	102	102

This table depicts the result of the questions that describe the feeling of security regarding data and Office 2.0 solutions of a user. The general feeling of security was determined with Q3.

When reviewing the table above, the factors fear of data non-confidentiality, the fear of losing data, the attitude towards whether sensible data that is worth protecting will be stored and the professionalism including trustworthiness can be identified. These factors in combination form the general feeling of security. The expected frequency of data non-confidentiality does not correlate with the general feeling of security. Thus data security seems to be a general principle for the user.

Whether sensible data will be stored or not is determined by the factors of general security, fear of cases of data non-confidentiality and the fear of losing data. Factors that do not influence it are professionalism and expected frequency of data non-confidentiality.

The professionalism and trustworthiness are determined by the same factors except the attitude towards whether to store sensible data on the service or not. The statement that data security is a principle can be reinforced.

Table 15. Correlations of productivity

Analysis of correlation									
		Q3	Q4	Q5	Q6	Q2	Q7	Q8	Q9
Q1	P. c.[a]	-.317	-.210	.291	.289	.111	.294	.002	-.008
	B. s.[b]	.001	.034	.003	.003	.265	.003	.986	.939
	N	102	102	102	102	102	102	102	102

[a]Pearson correlation
[b]Bilateral significance

Hypothesis 6. The questions Q1 (Do you feel more productive with this Office solution than with others/your current one?) and Q2 (Do you think that you could save time if merging documents created by different authors could be avoided?) show strong correlation Table 15.

The analysis of correlation shows that question 1 has no significant correlation with question 2. The hypotheses has to be rejected. This means that one core function of Google Docs, the possibility to edit the same document at the same time by different authors, does not add value to the productivity the user experiences.

Taking a closer look at the results of the analysis one can identify potential factors that correlate with the feeling of productivity. Q3 (Do you think that in comparison to stand-alone products you have to invest more energy for the same outcome?), Q4 (Do you think that in comparison to stand-alone products you have to invest more work for the same outcome?), Q5 (Do you think that in comparison to stand-alone products you get a higher quantitative output for the same input of work and energy?), Q6 (Do you think that in comparison to stand-alone products you get a higher output in terms of quality for the same input of work and energy?), Q7 (Do you think that this product will be extended in the future and that the new tools would lead to more productivity?), Q8 (The software forces me to perform tasks that are not related to my actual work), Q9 (My impression is that very little effort is involved in correcting mistakes).

3.4 Conclusion, Limitations and Further Research

This empirical survey reveals first insights into the usage of an Office 2.0 application. The results of the analyses can be interpreted to imply that users have doubts regarding data security when using Google Docs. For that reason providers of Office 2.0 software should offer better security and data confidentiality standards when it comes to storing data or documents in the cloud or transferring them over the internet. Due to the privacy and data confidentiality issues it can be assumed that especially for business purposes Google Docs is not eligible.

In terms of usability Office 2.0 solutions should have an easier and faster start-up. A short introduction should be provided in order to highlight the core functions as well as the main advantages compared to stand-alone solutions. Video-based trainings or tutorials could be another solution to make the advantages more popular. In the context of enterprises a help-desk to support user should be implemented. In general stand-alone applications are the benchmark.

The study shows that one core advantage of Office 2.0 applications, the possibility that different authors work at the same time on one document, does not add value to the felt productivity from the users' point of view. This result is quite interesting and needs further investigation.

In order to be more accepted and used that the Office 2.0 software Google Docs needs a thorough revision in order to be a serious competitor of stand-alone office applications.

The limitations of the study at hand lie in the fact that 102 valid answers were gathered. Furthermore the selected target group do not represent the current workforce. It was the intention of the study to take a closer look at the workforce of tomorrow. Enlarging the target group in order to get results from participants with different demographic backgrounds is another possibility to create more knowledge and deeper insights. The results of this study are only valid until the next major software upgrade by Google and ought to be renewed in order to keep up with the progress of the software. Furthermore the experiment was not too complex and the participants were mainly IT-savvy.

This study is the first part of a series of empirical studies which have the aim to investigate different aspects of Office 2.0 tools. The next step is to finish the already started survey with Microsoft 365. This survey is currently conducted with participants with the same demographic characteristics. The publication of these findings should provide a good basis for further research and for long-term results. This might be helpful to identify different fields of application in which this kind of tools is applicable.

References

1. O'Reilly, T.: What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software (2005)
2. Reinke, H., Bernhard, S., Björn, B., et al.: Microsoft Office 365 - einfach online zusammenarbeiten: [aktuell, schnell, sicher: neueste Techniken für die Kommunikation und den Datenaustausch in Unternehmen effizient einsetzen und nutzen]. Microsoft Press Deutschland, Unterschleißheim (2011)
3. Shari, S.C.S., Eldon, Y.L., Ya-Ling, W., et al.: Understanding web 2.0 service models: a knowledge-creating perspective. Inf. Manage. **48**, 178–184 (2011)
4. Ziyu, C., Haining, A.: The building of digital archives personalized service website based on web 2.0. Phys. Procedia **25**, 2096–2102 (2012). doi:10.1016/j.phpro.2012.03.355
5. Pieri, M., Diamantini, D.: An E-learning web 2.0 experience. Procedia – Soc. Behav. Sci. **116**, 1217–1221 (2014). doi:10.1016/j.sbspro.2014.01.371
6. Berthon, P.R., Pitt, L.F., Plangger, K., et al.: Marketing meets Web 2.0, social media, and creative consumers: implications for international marketing strategy. Bus. Horiz. **55**(3), 261–271 (2012). doi:10.1016/j.bushor.2012.01.007
7. Mell, P.M., Grance, T.: SP 800-145. The NIST Definition of Cloud Computing. National Institute of Standards & Technology, Gaithersburg, MD, United States (2011)
8. Murah, M.Z.: Teaching and learning cloud computing. Procedia – Soc. Behav. Sci. **59**, 157–163 (2012). doi:10.1016/j.sbspro.2012.09.260

9. Shahzad, F.: State-of-the-art survey on cloud computing security challenges, approaches and solutions. Procedia Comput. Sci. **37**, 357–362 (2014). doi:10.1016/j.procs.2014.08.053
10. Che, J., Duan, Y., Zhang, T., et al.: Study on the security models and strategies of cloud computing. Procedia Eng. **23**, 586–593 (2011). doi:10.1016/j.proeng.2011.11.2551
11. Sun, D., Chang, G., Sun, L., et al.: Surveying and analyzing security, privacy and trust issues in cloud computing environments. Procedia Eng. **15**, 2852–2856 (2011). doi:10.1016/j.proeng.2011.08.537
12. Pardeshi, V.H.: Cloud computing for higher education institutes: architecture, strategy and recommendations for effective adaptation. Procedia Econ. Finan. **11**, 589–599 (2014)
13. Rahmani, H., Sundararajan, E., Ali, Z., et al. (2013) Encryption as a Service (EaaS) as a Solution for Cryptography in Cloud. Procedia Technology 11
14. Avram, M.G.: Advantages and Challenges of Adopting Cloud Computing from an Enterprise Perspective. Procedia Technol. **12**, 529–534 (2014). doi:10.1016/j.protcy.2013.12.525
15. Plummer, Q.: Google rolls out high-octane drive for work offering (2014)http://www.techtimes.com/articles/9340/20140627/peek-google-drive-work-update.htm. Accessed 16 Jan 2015
16. Google: Google terms of service – privacy & terms – Google (2014) https://www.google.com/intl/en/policies/terms/. Accessed 16 Jan 2015
17. Crano, W.D., Brewer, M.B.: Principles and Methods of Social Research, 2nd edn. Lawrence Erlbaum Assoc., Mahwah and NJ (2002)
18. Willumeit, H., Hamborg, K.C., Gediga, G.: IsoMetricsS: Fragebogen zur Evalutation von graphischen Benutzungsschnittstellen (1997)
19. Willumeit, H., Hamborg, K.C., Gediga, G.: IsoMetricsL: Questionnaire for the evaluation of graphical user interfaces based on ISO 9241/10 (1998)

Component Database Management System Design Using Version Control

Fan Yang[(✉)], Zhenghong Dong, and Shucai Zhang

Academy of Equipment, Beijing 101416, China
sarah0824@hotmail.com, dzh.bj@163.com,
zhangshc@sina.com

Abstract. This paper builds up a database management system to manage algorithm components, where a component is consisted of its method to describe the basic information of the component and its software including the source code and executable program. Software and its method could be revised from time to time during the history of using. User may want to check the old version of a component. So it is necessary to keep the record of different versions of components in the database management system, as well as building up their correlations in between. Also, an algorithm component may have different versions when applied to different applications, but the core algorithms are the same. So, it is also very necessary to manage different versions of the same algorithm. However, no existing work has done similar work to manage the data in a database system. This paper applies version control system to manage the changes of different versions of methods and software, so that users can easily check the information of previous versions, or recover a component to a previous version. In the end, this paper also provides a prototype of the proposed database system.

Keywords: Version control · Database management system · Component management

1 Introduction

Version control systems contain large amounts of historical information that can give deep insight into the evolution of a software project [1], including the code age, number of bug fixes [2], structure information [3], which extracts meta-information from the software repositories [4, 5]. Subversion [6] is one kind of version control method which helps coordinate team development for large complex projects. They permit developers to work simultaneously on the same software system while ensuring that their modifications do not interfere with work done by other team members.

Version control is mainly used to keep the changed record of server files [7]. User can download server file to local, revise it, and commit the copy to server again. If the original file has been revised by other user, the version ID in the server is different from the local one. User has to combine the two versions into one and commit it to the server. In this way, user has to record the revised information, so that it is easy for user to look up for the changed information in future work. However, this kind of technology is only used for keeping the record of programming information, but not well used to database management, where big data is required to keeping the changed information.

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 404–413, 2015.
DOI: 10.1007/978-3-319-21009-4_31

This paper works on building up a database management system to manage algorithm components. A component is consisted of a method and software. The method contains the basic information of an algorithm, such as method name, creator, method classification, parameters, etc. And the software contains the executable program and source code of the algorithm. Users always update the old version of methods or software. However, current work of Database management only saves and manages data. Once the data has been changed, the new data simply replaced the old one. They did not keep the changed information. And some work may keep the old version of the data [8, 9], but did not keep the associations in between. It is not easy for user to find their relations. In fact, people would like to know the previous versions, and see where the files have been changed. Also, an algorithm component may have different versions when applied to different applications, but the core algorithms are the same. So, it is also very necessary to manage different versions of the same algorithm. In this case, this paper has involved Version Control in the proposed method to keep different versions of components. And also, users can recovery components from previous versions.

This paper is organized as follows. Section 2 discusses related works. Section 3 describes the main system. Section 4 provides the prototype design of the system. And Sect. 5 concludes this paper.

2 Related Work

2.1 Database Management System

In reference [1], the database management system stores log files of spatial data. They use Oracle database triggers to monitor database events and record the triggered information in the database log table. They also design an update log data. However, only updated type can be recorded, there is still no record of the old version.

Reference [10] also applies UML to build database management system, and solve the problem of realizing real time interactive and updating and maintaining via Internet. However, they still did not consider the problem of keeping the record of changed data, but simply updated them.

2.2 Version Control

Subversion is an open source system of version control. Compared with RCS, CVS, it applies the branch management system. Subversion keeps files in the center version repository, which is like a file server. The difference is, it can keep the record of updating information of files and contents, and user can recover data to previous version and view the data changed details. Considering that CVS cannot control the contents, but only files, while Subversion can control both contents and files, in the meantime, system requires the variety of client server. SVN [6] also can well support different systems, and the software is open source.

SVN manages changed files and contents, stores all files as binary files, and applies effective algorithm of comparing binary difference to compute the changes of different versions. In the meantime, it is a time machine, which keeps the record of every change of files and contents, for example, add, delete, or reorder files, etc. SVN also allows user to recover the data of previous versions, view the changed record of data. The key is to deal with the problem of version conflict, so SVN adopts the version control program of "lock – revise - unlock". SVN can support both windows and Linux operation system, and has good stability and safety on different operation systems.

3 Component Database Management System Design

This section describes the component database management system, including the basic function of the database management system, and especially, we discuss how we apply version control to this system in details.

3.1 System Function

Figure 1 shows the system function framework in branches. Component database management system mainly has three functions, synthesis management, method database management, and software database management. And each function has several different sub-functions.

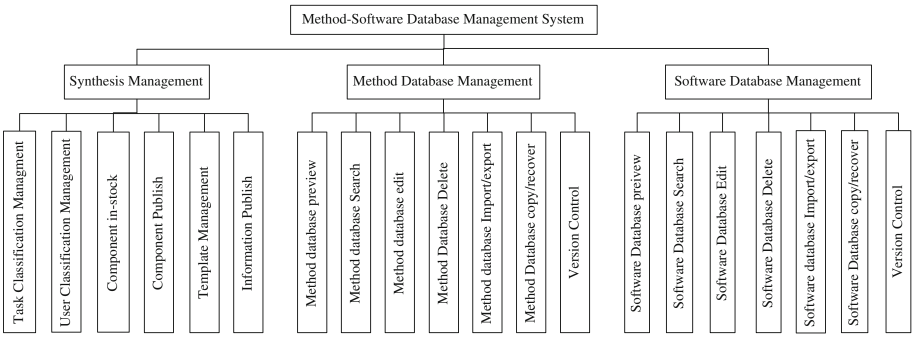


Fig. 1. System function framework

3.1.1 Synthesis Function

In the synthesis function, there are 6 sub-functions, which are task classification management, user classification management, component in-stock, component publish, template management, and information publish.

(1) **Task classification management** can add, delete, revise, search, and make statistic for tasks.
(2) **User classification management** can add, delete, revise, and search users. And set user to one of the 4 classifications: system management staff, servicer, component developer, and task application staff.
(3) **Component in-stock**: administrator check if the components meet up the test criteria, which includes Component contents are complete: equation, code, file,

compiling results, and test report are all completed. (b) Component contents (i.e. equation, code, and compiling results) should be correspondent to each other. (c) Component test samples should be correct and valid, and the component should pass the test.

(4) **Component publish**: if the component meets up the publish criteria, administrator can publish the component, and select which user classification to publish. From the component save, component in-stock registration, component in-stock, to component publish, component state in different f stages are undetermined, in-stock verification, in-stocked, and published.
(5) **Template management**: the system has 3 templates, which are code template, file template, and test report template. Also, administrator can add, delete, and revise self-defined template.
(6) **Information publish**: administrator can publish system announcement to a certain group of user, such as the in-stock information of new component, publish information, or server maintenance information, etc.

3.1.2 Method-Software Database Management

For the management of method database and software database, they share the same functions, so we introduce their functions together. Method-software database management mainly provides the basic management of database, including preview, search, edit, delete, import/export, copy and recovery, and version control.

(1) **Method-software database preview**: user can preview 2 different types of components,

   – If the user has authority and the component has been published.
   – If the component is developed by the user, and the component is state is "undetermined", "in-stock verification", or "in-stocked".
   Also, user can check the detailed information of the component, including equation, code, file, and test report.

(2) **Method-software database search**: user can select search type (including basic material, component parameter, component file, and component code), input search keywords, or select search method (such as selected by creator, component type, create time, keywords, etc.)
(3) **Method-software database edit**: administrator can edit the basic information of components. Once the administrator completes edit operation, and submit it to the server, the version number of the component will increase automatically.
(4) **Method-software database delete**: administrator can carry on the delete operation. And once the component is deleted, its related content would be deleted as well.
(5) **Method-software import/export**: user can select component for export, and export component information, equation, code, and file in the database and FTP server to local in the unified content structure. And also, user can select the local file, and import it to the database and FTP server in the unified content structure. The import/export function can be well applied to the update and synchronous between portable computer and database server.

(6) **Method-software database copy/recovery**: database copy operation can copy all data file in the database to FTP server, and set the basic information, such as name, time, copy reason, etc. Recovery operation supports user to recover data from selected multiple data.
(7) **Version control**: administrator can check the component version information, including the latest version, published version, and correspondent method version and software version. As a component is consisted of method and software, once the method or software version is updated, the component version is updated as well. If the basic information of the component is updated, the correspondent method file is updated, so the method version is updated, and the component version is updated. If the component parameter, equation have been updated, then the method file and code will be both updated, so both method version and software version are updated, then the component version is updated.

Version control is the key technical in this paper, so we discuss it in a separate sub-Sect. 3.2.

3.2 Version Control

Version control can effectively manage component version information in the database. The table structure of database includes component definition information, revise people, revise time, and version number, etc,. Version control supports the function of component version search, view and edit.

In this paper, version control tracks and control the update of component, including source code, file, and webpage. Server should build up central repository, and the client should support the basic function of version control.

3.2.1 Repository

Repository is a centralization system of storage and data sharing. Repository stores information in the structure of a tree and its branches. Any client can access to the repository to read and write this information.

3.2.2 The Basic Function of Version Control

Version control has 6 basic functions, which are sign out files or contents, submit files or contents, check out, update, version information view, and difference.

Sign out files or contents can copy the latest version from repository to the workspace. When sign out the contents, all files and contents under the contents will be signed out.

Submit files or contents: the submission means all updates have been saved to the repository.

Check out: check the repository will create the copy of branches in the local computer. The copies contain the latest version of the repository.

Update: repository can number different version by time, user can view any submitted version, and make any version in the work space go back to another previous version.

Version information view: user can check version information including path, version number, user name, etc.

Discrepancy: user can check the differences between two versions by commands, and compare the changes.

3.2.3 Solution

The current version control software includes CVS, SVN, etc. SVN can change files and contents, save files in binary format, and user binary difference algorithm to compute the changes between different versions. In the meantime, it is a time machine, so it changes with the changes of files and contents. SVN also allows user to recover previous version, or check the changed history of data. SVN adopts "lock – revise - unlock" strategy to solve the problem of version conflict. And also, SVN supports different operation system including Windows and Linux, and stay stable.

Considering that CVS has no way to control content version, but only file version, while SVN can control both file version and content version. In the meantime, system requires that client should be variety, and SVN has good support for different operation systems, such as Windows, Linux, and MAC. And also, the software is open source, so this paper chooses SVN to control version.

Version database is the center repository of data, which is the key part of SVN, so version database saves information in a tree structure of typical files and contents. Any client can access to SVN version database, read and write these files. Client shares information with the others by writing data, and obtain information by reading data.

3.2.4 Version Control Flow

The structure of system version control has 3 layers (show as Fig. 2), which are client/browser, Web server/SVN client, and database server/SVN server. Through browser, user uses version control command to upload components to SVN client, which is on the Web server. Next, user uses SVN command to publish components to SVN server, which is on the database server. Both SVN client and server support for different operation systems.

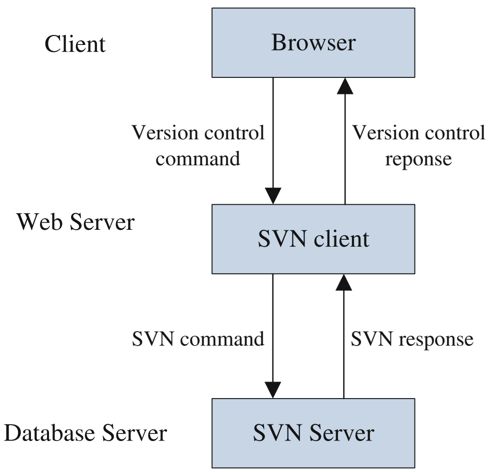


Fig. 2. system version control flow chart

4 Prototype Design

According to the system framework built above, we have provided the system prototype design as Figs. (3, 4, 5, 6 and 7). Figure 3 shows the synthesis management platform, where administrator can manage component type, user type, user authority, template, etc. Also, user can search version information of all components according to their create date. Figure 4 shows the component information publish platform, where

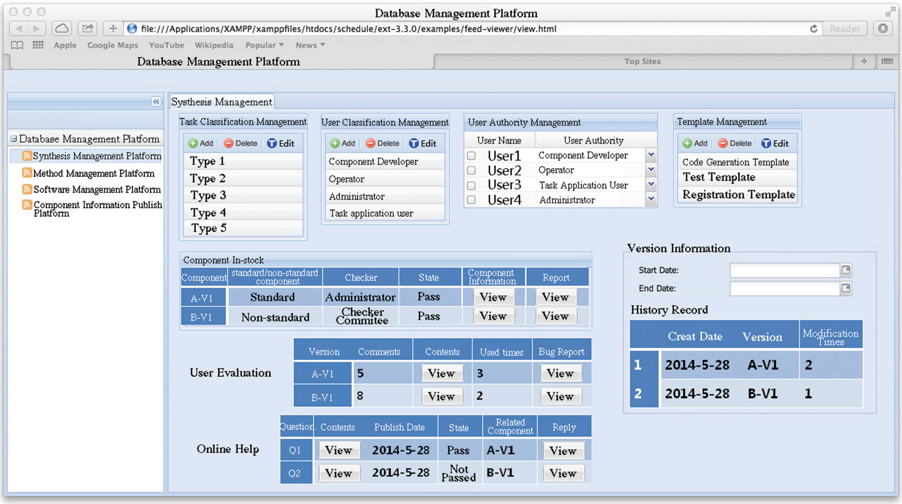


Fig. 3. Synthesis management platform

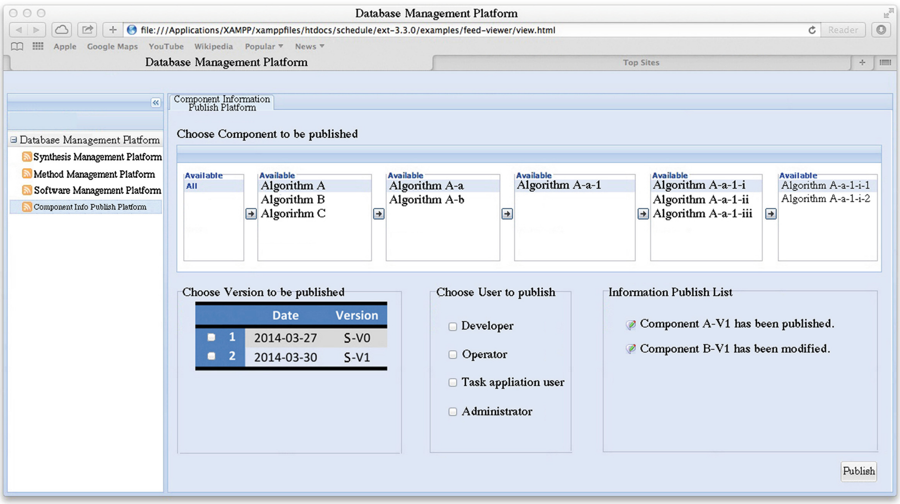


Fig. 4. Component information publish platform

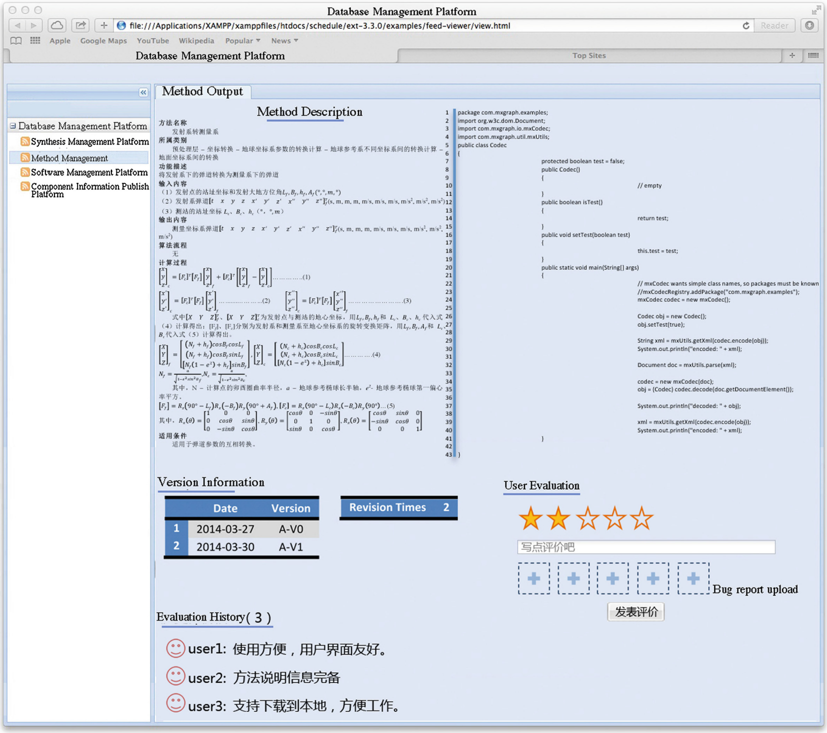


Fig. 5. An exported method

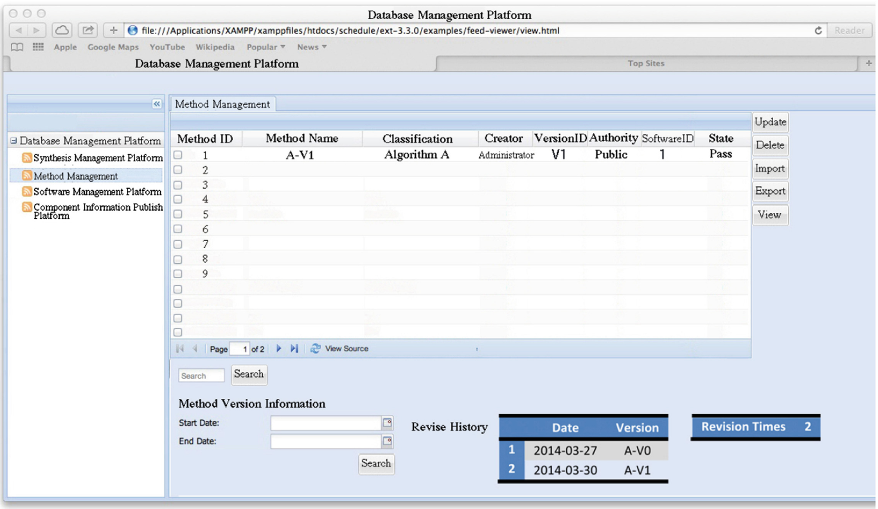


Fig. 6. Method management platform

administrator can choose which version of the component to publish, whom to publish. Also, news related any component that has been create, modified, or delete is provided in the information publish list. Figure 5 shows an exported method, which displays the method description, code, version information of a component. User can check the detailed information of each component. Different versions of this method are all given here, so that user can select the one he needed. Figures 6 and 7 shows the method and software database management platform. Figure 6 shows the method management platform, where all methods and their contents are saved, administrator can add, delete, update, import, export, search, or view them. User can find all versions of software that related to the method. Figure 7 shows the software management platform, where all software and its content are saved, administrator can add, delete, update, import, export, search, or view them. Also, user can find all versions of method that related to this software.

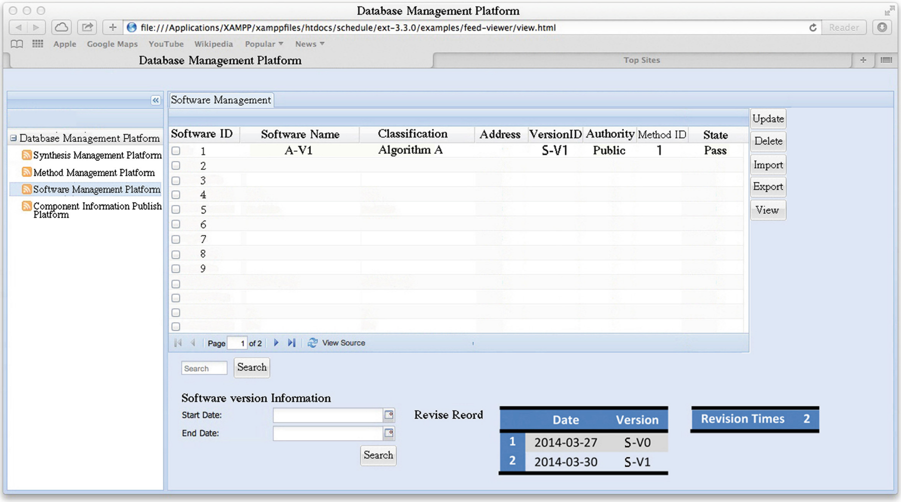


Fig. 7. Software management platform

5 Conclusion

This paper builds up a component database management system, where the components are consisted of methods and software. A method contains the basic information of an algorithm, such as method name, creator, method classification, parameters, etc. And the software contains the executable program and source code of the algorithm. Software and its method could be revised from time to time. It is necessary to manage different versions of components in the database management system. In order to solve this problem, this paper applies version control system to control the changes of different versions, so that different versions of component have been kept in the database, as well as their correspondence in between. Users can check any previous version of component, and also recovery a component from previous versions. In the end, this paper provides a prototype to display how we design the database system.

References

1. Cherait, H., Bounour, N.: Toward a version control system for aspect oriented software. In: Bellatreche, L., Mota Pinto, F. (eds.) MEDI 2011. LNCS, vol. 6918, pp. 110–121. Springer, Heidelberg (2011)
2. Christopher, M.B.: Taylor and Malcolm Munro: Revision towers. In: Workshop on Visualizing Software for Understanding and Analysis VISSSOFT, pp. 43–50. IEEE Computer Society Press, Paris (2002)
3. Gall, H., Jazayeri, M., Riva, C.: Visualizing software release histories: the use of color and third dimension. In: International Conference on Software Maintenance (ICSM), pp. 99–108. IEEE Computer Society, Oxford (1999)
4. Hindle, A.J.: SCQL: A Formal Model and a Query Language for Source Control Repositories. A Master thesis. University of Victoria (2005)
5. Hassan, A.E., Holt, R.C.: The Chaos of software development. In: International Workshop on Principles of Software Evolution (IWPSE), Helsinki, Finland (2003)
6. Nagel, W.: Subversion Version Control: Using the Subversion Version Control System in Development Projects. Prentice Hall Professional Technical Reference (2005)
7. Huang, Y., Lai, J.-Q, Mao, X.-C., Fan, : Modeling of Web Geological Database Management Based on UML. In: International Conference on Computation Intelligence and Software Engineering, pp. 1–4 (2010)
8. Vassilakis, C., et al.: A Comparative Study of Temporal DBMS Architectures. In: Proceedings of DEXA 96 workshop, Zurich (1996)
9. Sotiropoulou, A., et al.: Temporal Extension to ODMG. In: Proceedings of the 3rd Biennial World Conference on Integrated Design and Process Technology, and Applications of Database Technology (IADT), vol. 2, pp. 304–311 (1998)
10. He, S., Liu, G., He, Z., Weng, Z.: Design and implementation of log management module in three-dimensional spatial database management system. In: International Conference on Geoinformatics, pp.1–5 (2010)

Multi-Column Query Method Research and Optimization on HBase

Bo Shen[1(✉)], Han-Chieh Chao[2], and Weijia Xu[1]

[1] School of Electronic and Information Engineering, Key Laboratory of Communication and Information Systems, Beijing Municipal Commission of Education, Beijing Jiaotong University, Beijing, China
bshen@bjtu.edu.cn

[2] Department of Electronic Engineering, National Ilan University, Yilan, Taiwan

Abstract. This paper focuses on analyzing how to improve the HBase non RowKey's query and designing an approach to improve its performance. Analyzing and summarizing the advantages and disadvantages of index technology applied in typical application scenarios, we design a secondary index approach for HBase based on the pre-partition. Through experiments, we found this method can effectively improve the query performance of HBase on non RowKey column, have little effect on the performance of the original data writing and reduce data redundancy. Compared with other methods, this approach has certain performance advantages.

Keywords: Hbase · Rowkey · Index · Query performance

1 Introduction

With the development of the Internet industry, traditional relational databases can't satisfy the needs of users, so Nosql begins to play an increasingly important role [1].

In Nosql family, HBase is an excellent product. The HBase merely establishes B + Tree index structure for RowKey, while it uses the full table scan for the non-RowKey inquiries, leading efficiency very low [2].

Based on the general demand, developers hope that HBase can maintain a high performance advantages, meanwhile, it is able to support the complex conditions query [3].

Given the above, focusing on Hbase, this article designed index structure suiting for Hbase. Respectively, we detailed designed the index of the logical layer and the storage layer, and optimized the data consistency after the index established. We make possible HBase for multiple columns with high performance in query and low-invasive without much compromising data consistency and low redundancy of the original data.

2 HBase Principle

HBase is a distributed database, managing clusters by ZooKeeper. On the architectural level it can be divided into Master and multiple RegionServer [4].

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 414–421, 2015.
DOI: 10.1007/978-3-319-21009-4_32

RegionServer is the core module in Hbase, which corresponds to a node in the cluster, and one RegionServer is responsible for managing a lot of Regions. HBase provides three Split methods [5], which are pre-partition, automatic and manual partition enforcing to meet different business demands.

There are two special tables in HBase: -ROOT- and.META. They work together with ZooKeeper to achieve query mechanism. -ROOT- is used to record the information of Region from.META. table, while it can only be stored in one Region. In contrast to -ROOT-, .META. can be separately stored in more than one Region, used to record the Region information in user table, The ZooKeeper is available to recording the location of the -ROOT- table saved [6].

3 Research on HBase Multi-Column Query Method

HBase improves the capability of reading and writing, at the same time it ignores efficient getting feature under the multi-column. Based on the different structure, there are three methods to solve it which are secondary index, double index and index based on the linearization technology.

The secondary index has been widely used in the relational database, whose two main methods are IHBase and CCIndex. Implementation on the secondary index by building the index table, not setting up a complex hierarchy of the index table, makes it easier to be created and maintained. However, the method creating an index table to provide index mechanism will proportionally increase the data storage and data redundancy.

The index based on linearization technique such as MD-HBase building the KD-Tree data structure and realizing dimensionality reduction through the Z-order solves the problem of multi-dimensional data query and has a good performance in the processing of space data. But it's much difficult and complex to map the text data to a multidimensional vector space and divide by KD-Tree in dealing with text data.

Double index method like EMINC coordinates global index and local index. The global index builds the R-Tree data structure, and the local index uses a KD-Tree data structure. We locate the node where data is by the global index, and then detects the data required through the local index. Such process of pruning reduces the extra query cost and avoids unnecessary inquiries; In terms of the insertion, it makes much easier to put the similar data together. But during the local index updating, it cost a lot for maintenance of global index.

Given these problems, the method in this paper will refer to the idea of secondary index method and design it better.

4 Design of the Pre-Split Secondary Index Approach

This part presents a secondary index method based on pre-partition. The design of its index logic layer and the data storage layer will be introduced. The method will be optimized and data consistency will be ensured [7].

4.1 1The Design of the Index Logic Layer

The design of logical index layer consists of single-column index construction and multi-column index construction. The single-column index exchanges the TRowKey of the main table and the *TValue* of the index row and stores in the index table (Index-Table). Here, the *IRowKey* of the index table is designed like this:

$$IRowKey = CF : Qualifier + TValue + TRowKey$$

CF is the name of index column family. *Qualifier* is the name of the index column. The design can ensure uniqueness of the *IRowKey*. In order to facilitate updating and maintaining the index table, a column named status is added to be a marker of updating and deleting the index. In order to save storage space, the former 16 KB of the *TValue* works as the index marker which can be modified in the client.

The procedure of multi-column index is similar to the single-column index, but they have great differences in performance. Under a two-column index built, the *IRowKey* for inverted index method is designed as follow:

$$IRowKey = CF1 : Qualifier1 + CF2 : Qualifier2 + TValue1 + TValue2 + TRowKey$$

Here, *CF1* is the name of Column Family for the original data of the first column. *Qualifier1* is the name of Column for the original data of the first column index. *CF2* is the name of ColumnFamily for the second column index original data. *Qualifier2* is the name of the Column for the second column index original data. Because the query process of the HBase comply with the principles of left most prefix match, *CF1: Qualifier1* is listed as the most frequent query column.

4.2 The Design of Index Storage Layer

HBase realizes the communication among Client, Master and RegionSever by RPC mechanism [8]. But this protocol may bring damage because of traffic delays. Getting data across Regions will bring some RPC operation delay. To minimize such RPC process, the method designed controls the HBae data fragment and makes the main table data and index table data be split in the same location, which ensures that the main table data and index table data are split into the same Region. So we can launch two queries straightly without returning to the client reducing process of RPC.

In order that data and corresponding index table data are sliced to the same Region, in this paper, we use the Region Observer in Coprocessor to intercept the writing operation but not the HBase automatic partition strategy. The index table data will be written into MemStore according to the data in main table. The preFlush method is provided to ensure the correspondence of the main table data and the index table data. Meanwhile, in order that the main table data and the index table data are in the same Region, this paper disuses the HBase itself fragmentation and takes the pre-partitioning algorithm: some Region for storing data table are pre-set and writing operation is followed. The algorithm will be described as below.

4.3 The Design of Pre-Split Algorithm

In this paper, for avoiding the problem of hotspot without pre-split, we adopt pre-split algorithm to improve the design of index storage layer. Algorithm implementation process is shown in Fig. 1.

The low number of Region would influence on performance of concurrency, the high would bring inconvenience in system management and excessive memory usage. In this paper, the algorithm gets the reasonable number of Region by combining recommended system values with subjective foresting values.

The average number of Region for a table created about from half to one hour can be calculated by the following formula:

$$R_{\mathrm{avg}} = \sum_{\mathrm{i}=1}^{N} (R_{\mathrm{i}} * G_{\mathrm{i}}) \Big/ \sum_{\mathrm{i}=1}^{\mathrm{N}} G_{\mathrm{i}} \tag{1}$$

Where G_i is the weighted value of each R_i determined by the time difference between S and S_i. Gi values {0.1(|S-Si| < 1800),0.2(|S-Si| > 3600),0.7(1800 < |S-Si| < 3600)}. Then comparing recommended system values with subjective foresting values, the larger one is the number of Region N.

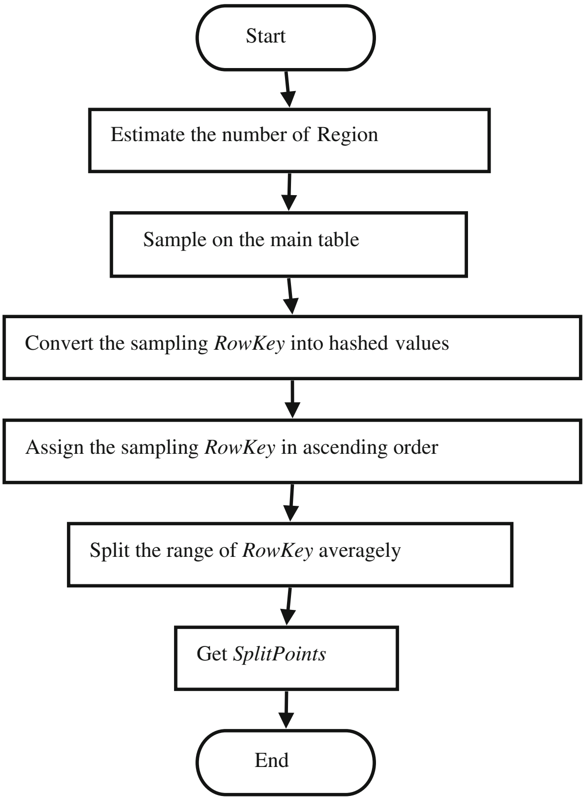


Fig. 1. Pre-split Algorithm

Next, sampling on the main table we can get sets RowKeys[], then we convert RowKeys[] into hashed values for getting MD5_RowKeys[] by MD5 algorithm, and assign the sampling RowKeys in ascending order to get ASC_MD5_RowKeys[]. Now, we can estimate the range of the Rowkey of tables. According to the number of Region N, averagely splitting the whole sets, we get N-1 SplitPoints.

4.4 Data Consistency

In HBase, if we don't create secondary index table, we could only ensure the correctness of data written in one table. However, faced with creating secondary index table, we need ensure both correctness and consistency of writing data between main table and index table. That inevitably brings the problem of data consistency. In this paper, to solve this problem, we use local transaction of HBase [9].

There are two situations which need ensure data consistency in this paper: one is the main table writing, the other is the data of main table updating. In the main table writing, we write data of the main table and data of index table into the same local transaction. That is, we think data of the main table writing and data of index table writing as a whole operation which writes into WAL log and makes persistence; for the data of main table updating, we package three operations which include updating data of the main table, marking old data of index table invalid and adding new data of index table into a local transaction. That is, we take the main table writing and two index tables writing as a whole, written into WAL log and made persistence.

So, introducing transaction operation, but for the times of writing WAL log on the whole operation, we transform twice or three times into current once. Because of taking twice and three times operation as a whole operation, faced with introducing transaction, we can still improve the performance of data writing, at the same time, ensure operating data latency not significantly increased.

5 The Performance of the Simulation and Comparison

This article selected the wikipedia data, through four computers to build distributed cluster environment. To test performance of method this paper provided, we performed four experiments about single-column query, multiple-column query, data writing and data redundancy, and compared with CCIndex, IHBase and indexless HBase, which also used the idea of secondary indexes. Finally, we analyze and prove the experimental results from the theory.

5.1 Single-Column Query Performance

We inquired the data for wiki: anchor = Etymology in 0.1 million,1milliion and 5 million data, and separately get the query time. In our experiment, each method had the same original data. The experiment results shown in Table 1:

Table 1. Single-column query results

Test cases	***0.1 million data***	***1 million data***	***5 million data***
Indexless	0.57 s	0.905 s	1.118 s
Our method	0.046 s	0.073 s	0.092 s
IHBase	0.083 s	0.114 s	0.125 s
CCIndex	0.043 s	0.070 s	0.091 s

5.2 Multiple-Column Query Performance

We queried wiki: anchor and wiki: link simultaneously in 0.1 million,1milliion and 5 million data, separately get the query time. In our experiment, each method had the same original data. The experiment results shown in Table 2:

Table 2. Multiple-column query performance results

Test cases	***0.1 million data***	***1 million data***	***5 million data***
Indexless	0.75 s	1.046 s	1.373 s
Our method	0.063 s	0.084 s	0.099 s
IHBase	0.097 s	0.133 s	0.167 s
CCIndex	0.078 s	0.106 s	0.128 s

5.3 Data-Writing Performance

Using four methods to set 1.5 GB original data, we specified column Anchor as index column and get the writing time separately for each method. In experiment, each method took the same data. The experiment results shown in Table 3:

Table 3. Data-writing performance results

Test cases	***Total throughput***	***Total time***
Indexless	89.82 MB/s	17.11 s
Our method	84.66 MB/s	19.44 s
IHBase	68.53 MB/s	24.81 s
CCIndex	63.91 MB/s	28.05 s

5.4 Data Redundancy

We wrote 1 million data respectively by the method, IHBase and CCIndex. Then, we build single-column index about column Anchor and multi-column index about column Anchor and Link, and get the total storage capacity each method occupied. At the same time, writing 1 million data by the indexless HBase as a redundancy comparison.In our experiment, each method took the same original data. The experiment results shown as follow (Table 4).

Table 4. Data redundancy results 1

Test cases	***The total storage capacity***
HBase	89 MB

From the results, we can conclude that the method has certain help to improve the query performance of HBase based on non RowKey and less impact on writing performance of original HBase. This fully shows the two advantages of this method: Reducing the times of communication about RPC in query process and improving the performance of HBase based on non RowKey. To avoid operating time delay regional fragmentation bringing, we adopt pre-split mechanism, mitigating the impact on writing performance for adding indexes to HBase Table 5).

Table 5. Data redundancy results 2

Test cases	***The total storage capacity***	***Data redundancy***	***Redundancy rate***
Our method(single column)	129 MB	40 MB	44.94 %
IHBase(single column)	135 MB	46 MB	51.68 %
CCIndex(single column)	227 MB	138 MB	155.05 %
Our method (multi-column)	195 MB	106 MB	119.10 %
IHBase(multi-column)	207 MB	118 MB	132.58 %
CCIndex(multi-column)	323 MB	234 MB	262.92 %

6 Conclusion

In this paper, analyzing the existing three types of index structure, and combining with the characteristics of HBase, we design a new secondary index method based on pre-split through secondary indexes design idea whose logic layer and storage layer has been designed in detail. Via performance simulation and comparison of IHBase, CCIndex and no index of HBase, we can find that the secondary index method based on the pre-split not only improve the HBase based on non RowKey query performance, also mitigates the effect on the properties of the original HBase writing. Compared with IHBase, CCIndex, the method has a good performance in query, writing and data redundancy. As a whole, it has certain performance advantages.

Acknowledgements. This work has been supported by the National Natural Science Foundation of China under Grant 61172072, 61271308, and Beijing Natural Science Foundation under Grant 4112045, and the Research Fund for the Doctor-al Program of Higher Education of China under Grant W11C100030, the Beijing Science and Technology Pro-gram under Grant Z121100000312024.

References

1. Bakshi, K.: Considerations for big data: Architecture and approach. In: Aerospace Conference, 2012 IEEE, pp. 1–7. IEEE (2012)
2. Wang, J.: Index and query processing technology research in the cloud computing system. HIT (2013)
3. Du, X.: Big data environment based on HBase distributed query optimization research. Computer CD Software And Application 8 (2014)
4. Vora, M.N.: Hadoop-HBase for large-scale data. In: 2011 IEEE International Conference on Computer Science and Network Technology (ICCSNT), pp. 601–605 (2011)
5. Liu, X.: The depth analysis of the HBase performance. Programmer **7**, 102–104 (2011)
6. Zhuo, H.: Real-time query system desigan and implementation based on HBase huge data, BUPT (2013)
7. Vashishtha, H., Stroulia, E.: Enhancing query support in hbase via an extended coprocessors framework[M]. Springer, Berlin Heidelberg (2011)
8. Yunzhi, G.: Application of Remote Procedure Call Protocol(RPC) in Heterogeneous Dual-Core Communication[J]. Fire Control Radar Technology (2013)
9. Dean, J., Ghemawat, S.: MapReduce: simplified data processing on large clusters[J]. Commun. ACM **51**(1), 107–113 (2008)

Social Networks and Mining Techniques

Introduction to Social Business Process Management

Mateja Kocbek(✉), Gregor Jošt, and Gregor Polančič

Faculty of Electrical Engineering and Computer Science, University of Maribor, Smetanova 17, 2000 Maribor, Slovenia
{mateja.kocbek,gregor.jost,gregor.polancic}@um.si

Abstract. Technologies and concepts of Web 2.0 are strongly present in everyday life and also in business environments, where they are increasingly involved into business activities which constitute business processes. Combining the management of Web 2.0 and business processes resulted in social BPM, which is one of the most promising fields in the software industry. Social BPM tries to create knowledge in business environments and manage it efficiently. In this contribution, some key features are presented. We particularly exposed strengths, weaknesses, opportunities and threats of social BPM.

Keywords: Social BPM · Business process management · Web 2.0 · Enterprise 2.0

1 Introduction

Priority of every organization or company is to increase operational efficiency, reduce costs, improve quality of services and better retain operational knowledge. This requires innovative and robust IT solutions, which have to be carefully designed and implemented. From here comes one of the major challenges of modern organizations, namely the integration of business processes with business data [1].

Many organizations are using business process management (BPM) as a key component in automating workflows, increasing standardization and improve performance. They are also combining BPM with service oriented architectures to ensure repeatability, extend business capabilities and support asset reuse [1].

The BPM approach is relatively accepted in organizations. It represents the key component of automation of workflows, standardization, and general improvement of business operations [2]. Because of increased interest, many methodologies and paradigms are included in BPM [3], for example: theory of organization management, computer science, mathematics, linguistics and philosophy. Thus, BPM is becoming an interdisciplinary theory in practice [3].

Thus it can be said that BPM includes a variety of different technologies, actors and science in practice [3]. One of the recently included technologies is Web 2.0, which is the focal point of this article.

Technologies Web 2.0 within enterprise environment are called Enterprise 2.0. This represents combination of social programing and communication technologies.

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 425–437, 2015.
DOI: 10.1007/978-3-319-21009-4_33

Enterprise 2.0 was first mentioned in 2004 by Tim O'Reilly and Dale Dougherty [4]. Examples of Web 2.0 include social networking sites, blogs, wikis, video sharing, etc.

Social networks have shown a great potential for improved interpersonal communications, collaboration and group learning. Organizations are adopting these tools with focus toward communication, outreach and public relations. However, a very few organizations are leveraging the power of social Web to support creative problems solving and capture expert knowledge about complex processes [1], although a lot of them use social networks to promote their business and to reach potential users.

BPM and Web 2.0 can solve many challenges of organizations, because social software supports various actors in producing user-generated content, developing and maintaining social relationship as well as establishing computer-mediated interaction and collaboration. These capabilities of social software provide manifold possibilities for more effective and flexible design of business processes [5]. But it is necessary to stress that some experts have exposed counter-arguments against social BPM. The problem arises because of interactions between people, since users generally expect instant gratification, rich user experiences, rapid access to information and constant connectivity, yet are not willing to cooperate for any price. Thus, organizations are faced with a challenge of modernizing their business processes and still attract participants [2].

In subsequent section we presented the main concepts which constitute social BPM: BPM, Web 2.0 and social networks. In Sect. 3, a brief introduction to social BPM is made by describing its development, examples an also a comparison with traditional BPM is given. In Sect. 4, a SWOT analysis is presented. Finally, Sect. 5 presents a brief summary.

2 Background

2.1 Business Process Management

Business Process Management (BPM) is an approach that helps organizations and companies to better manage their processes, gain higher productivity, reduce costs, or add value to their work in general [6, 7]. With this approach organizations are able to achieve their strategic goals in a more efficient manner. BPM can be viewed as a contact point between business and IT [8].

Furthermore, the continuous improvement of business operations can be performed through iterations and some profound changes of business processes [9]. This is in accordance with the life cycle, that includes design, engineering, enactment, monitoring and re-engineering phases [9, 10].

2.2 Web 2.0 (and Enterprise 2.0)

The second generation of Web, called Web 2.0, gained widespread interest of individual users as well as business world [11]. Web 2.0 is composed of a set of technologies, where the most prominent are considered to be blogs, mash-ups, peer-to-peer networking, RSS (Really Simple Syndication), social networking, web services and wikis [12]. One of the advantages of Web 2.0 is its interactive and collaborative orientation. Users are much more involved than they were in the first generation of Web and they mainly create the content [11, 13].

In conjunction with Web 2.0, another term has been coined, namely Enterprise 2.0. Enterprise 2.0 is basically an extension of Web 2.0. Both use the same tools or technologies and have a similar philosophy. The only major difference is in who is using these technologies; common users or enterprises. Web 2.0 is both designed for a common user and also primary used by one However, if an application of Web 2.0 (e.g. Facebook) is used for company's needs, this is considered as Enterprise 2.0 [14].

2.3 Social Networks

Social networks are one of the most used technologies of Web 2.0. They appeared in 2000 and the first better known example is MySpace, published in 2003. In 2004, Facebook was introduced, followed by other examples of Web 2.0 technologies, such as LinkedIn, Twitter, YouTube, Pinterest, Instagram, Snapchat, Flickr, WordPress, Blogger, Wikipedia, Second Life, etc. [15].

A reason for widespread use of social networks are user friendly pages. They are simple and demand only basic IT literacy. The great importance to the user is also the option of joining groups with common interests. The concept is similar to the concept of the association or other community. Most popular social networks are free to use whereas the service providers mostly earn with advertising.

Graph Theory method has been widely used in social network structure modelling. Harary [16] used directed graph model to describe the one-way relationship in the social network, and put forward the concept of centrality. Sociologists have studied many of properties of social networks. For example, six degrees of separation, small world phenomenon, scale-free, power law distribution and structural robustness, etc. The small-world phenomenon, popularly known as the six degrees of separation, has been experimentally studied by Milgram [16] in the field of social psychology.

Social networks can serve organizations to more easily access to their potential customers, and also to increase communication and collaboration within a company. However, there are some organizations, like hospitals, that have difficulties in adopting socially supported applications. There is also a concern that disseminating sensitive information on an enterprise social network can put them or their clients at risk. Other companies that are not exposed to such high business risk use social tools beneficially, while limiting unwanted exposure. More details about companies' adoption of social media and BPM are discussed in details in next chapter.

3 Social BPM

Most business process management activities are performed by method experts and IT developers according to given requirements [17]. But correctness, completeness and usefulness can be achieved only if relevant stakeholders participate in all phases of life cycle [18]. Thus, traditional BPM no longer fits to this view of processes. Its several shortcomings motivate the need for social approach. Social BPM, for researches and practitioners, addresses communication and collaboration support of all BPM stakeholders from a methodological and technological perspective, in order to improve business processes [19].

To create better and higher-quality process solutions that a single expert can achieve, so called “wisdom of crowds” was introduced [20]. This concept enables a large variety of internal and external actors to contribute their domain knowledge and method expertise [9]. Another crucial development of social BPM was also the rise of social networks, which can also be viewed as fundaments of social BPM.

There are many definitions of social BPMN. However, majority of the definitions mostly emphasize the role of collaboration during the process design stage of the BPM lifecycle and neglect how the social BPM engine could work during runtime [21]. One of such definitions classifies social BPM as “a methodology of bringing more and diverse voices into process improvement activities” [22].

Social BPM offers a platform that enables contribution from various users of the community during all the stages of the BPM lifecycle. This increased participation ensures the users actually follow the processes during enactment, which overcomes the model-reality divide limitation that has been identified in the traditional BPM system.

Traditional BPM has several weaknesses, which also caused the development of social BPM. Traditional BPM namely tends to follow a top-down decomposition strategy from business goals to business processes and activities [17].

Another major disadvantage of traditional BPM is that they too static. Today’s rapid markets changes should be reflected immediately. Traditional approaches turn out to be too time-consuming and costly, while social BPM (with its Web 2.0 components) can increase flexibility and responsiveness.

Through the development and the use of BPM, some issues and challenges appeared [23]. The first issue is *model-reality divide*, a demarcation between modelled and executed processes [24]. The issue often occurs when employees do not adopt “the model of business process”, but their “real” version. As a consequence on this, the *information pass-on threshold* was introduced by [24]. The issue addresses the users’ ideas, which frequently get lost or remain unrealized. This can explain why there is such a divide between model and real process [24].

The second issue is called the *lost innovation* and addresses the situation where the knowledge of stakeholders gets lost and potential improvements for company remain unrealized [24]. Another common issue is *lack of information fusion* that represents a lack of involvement of the participants in the modelling of business processes. The reason for this may be in companies’ internal structure (how the participants/employees are classified within the company) or in adopting modelling tools [24].

3.1 Comparison: Traditional and Social BPM

Traditional BPM has its roots in the concept of Scientific Management by F.W. Taylor, which aimed for efficient operation in a well-structured, stable, predictable economic environment. Since today’s business success in dynamic, global and competitive markets mainly depends on the ability to innovate, organizations have to rearrange themselves. Networks within and between organizations contribute to greater flexibility, versatility, and innovation in the organizations' ways of meeting customer needs and requirements [25, 26]. This obviously refers to products, services and processes. Consequently a continuous shift occurs from clearly structured and manageable organization forms with well-defined boundaries (information society) to highly dynamical, self-organizing network organizations with blurred boundaries (knowledge society) [19, 27].

The shift from organizations in the information society to organizations in a knowledge society is given in Table 1.

Table 1. From traditional to social BPM [19, 27, 28]

	Organizations in the information society	Organizations in the knowledge society
	Traditional BPM	Social BPM
Structure	Hierarchy	Network (Relationship)
Dynamics	Processes (highly repetitive, structured, predetermined, predictable processes)	Learning (case-based, less-structured, flexible, unpredictable processes)
Focus on	Decomposition and stability	Integration and dynamics
Measure	Towards high efficiency	Towards high effectiveness
Assets	Tangible	Tangible and intangible
Economic environment	Certainty, little change	Uncertainty, high dynamics
Values	Based on 'law of scarcity'	Based on 'law of abundance'
Jobs	Based on traditional skills	Based on intertwined skills and digital media literacy
Production	Oriented to mass	Oriented to mass customization
Products and services	Fordian	Value-added products and processes
Idea of organization	Efficient machine	Open, adaptive organism
Idea of man	Mechanistic (left-brained) people as cogs	Humanistic (right-brained) knowledge workers (creative, intelligent, thinking agents)
Organization of work	Central planning by experts ('white collar') for execution by others ('blue collar')	Decentralized, self-organized planning and execution by knowledge workers; Participation of stakeholders by social interaction and collaboration in communities

4 SWOT Analyses

The SWOT analysis was performed based on relevant literature that deals with one or more aspects of SWOT analysis. Our goal was to combine all aspects to gain joint perception of Social BPM. All aspects (Strengths, Weaknesses, Opportunities and Threads) are discussed further on.

4.1 Strengths

One of the main advantages of including social software in BPM first and foremost addresses the shortcomings of traditional BPM approach in all stages of BPM life cycle. Such shortcomings include issues with processes not being executed as modeled (also known as model reality divide), existing knowledge for improvements not being recognized or utilized (lost innovation) and participants not passing on improvement proposals (information pass-on threshold). To address such shortcomings, social BPM covers the typical BPM life cycle activities, mainly process design, implementation, deployment and execution [19].

So, while both traditional BPM and social software address the management of activities, the latter provides a variety of new features. Among those, the most prominent are represented in Table 2 [19, 24, 29]:

Table 2. Social BPM features

Feature	Description	Real-life examples
Self-identification	Participants can choose to contribute to the activity, deeming themselves competent	Wikis
Transparency	The results of the collaboration are available to everyone	Wikis, blogs
Signing	Participant signs the work activities when they are completed	Wikis, blogs
Open modification	If necessary, anyone can modify the work of other participants	Wikis
Logging	All activities are logged, consequentially providing a history of activities	Wikis, blogs
Discussion	Suggestions and comments regarding the work can be easily discussed	Blogs
Banning	If necessary, participants in the process can be banned.	Wikis, blogs
Tagging	Used to annotate business processes with process policies	Blogs

Since social software enhances the traditional BPM with the features represented in the table above, we gain opportunity to define business in different, innovative ways. This becomes prominent when the process demands a higher level of collaboration, which also requires a lot communication between the participants. Also, social BPM is best introduced to the processes, where services are not standardized.

Another main advantage of social BPM is also the integration of all stakeholders into the BPM life cycle [19], making the information available to all participants without additional efforts [29]. This is mainly because social BPM also supports the creation of *weak ties*, a concept that is crucial to improve the agility and innovations of the enterprise. While *strong ties* are imposed by the corporate hierarchy or team membership, weak ties represent the connections between individuals. This is also in align with the idea that the company must be opened in ordered to capture new ideas and create competitive products. Cooperation with many different participants is crucial and this is well supported with social BPM. Furthermore, by keeping company open, we implicitly achieve *wisdom of crowds*. This is an idea based on a premise that for any given problem a near-optimal solution can be found by combining as much inputs as possible. This translates to social BPM in a way that many participants collaborate in delivering a better solution [23].

After the process execution, social BPM also enables automatic communication and collaboration patterns (i.e. behavior of participants). This ensures the identification of good practices, which can be reused when working on future instances. Consequentially, the collective knowledge can be analyzed and may lead to model or implementation redesign [19].

4.2 Weaknesses

Since there is no official definition for social BPM, the concept can seem vague and is usually not fully understood. This might be one of the main reasons that many business owners hesitate to incorporate social BPM or do not give the concept a try [30]. Another important aspect is also perceived loss of management control over the processes, since an open platform suggests that many different participants work on a single process [31].

However, even when the advantages of social BPM are recognized, the acceptance of such features is questionable. While new social features do increase the effectiveness of collaboration and communication, they demand a certain amount of time for acceptance [24] and are usually not intuitive to use [29]. Also, since social BPM includes many participants, the release of new features must be strictly coordinated.

To better illustrate this weakness, in [32] authors explain that only a few percent of employees are using platform for cooperation between employees. The research showed that most of those who have adopted the platform, are already working with technology.

Furthermore, social software is considered to be unsecure. The risk of losing data or a security breach with Software-as-a-Service BPM tools is always present [31]. Access to information and registration process are commonly available to everyone, making social software not appropriate for managing critical business related data [29]. Some participants may also not be willing to make their comments or changes immediately visible on social platform [24] or even find the additional task of manually updating their work to be time-consuming [33].

Besides, as stated by Nathaniel Palmer, an executive director of the Workflow Management Coalition, there is currently no widely accepted corporate alternative to Facebook. All attempts at creating such an alternative quickly lose strength or momentum, not attracting enough participants to achieve meaningful and successful social BPM [30].

On the other hand, if an organization has not yet fully implemented its social business strategy, business owners may be subjected to many different social streams, such as Twitter, Facebook, LinkedIn, Chatter etc. Having too many different social streams and no mechanism to unify them at some point, the information becomes increasingly harder to find. Therefore, it would be ideally to have one unified social stream, but this has proven to be hard to achieve in real-world situations [33].

An open issue regarding social BPM is also whether if community-driven modelling and execution increases or actually decreases the model quality. Therefore, it is necessary to always measure the quality of models in a community-driven collaborative projects [29]. Besides, because we operate on an open platform, overheads of the contributions greatly increase [23]. But excessive information does not necessarily result in the correct information. We can be left with many unnecessary updates, incorrect information and, consequentially, wasted time and resources without any particular benefits [33].

On a more basic level, the main critiques of social BPM is that participants collaborate on the *finished* product, and not in the *development* stages. In this light, the correct usage of social software may completely replace the need for designers, since every participant will ultimately be a designer. As author argues, this will not be the BPM lifecycle using social software, but it will eliminate the BPM lifecycle altogether [34].

4.3 Opportunities

The following subsection discusses the opportunities of social BPM [24]. Social software enables rapid communication between participants. So, getting all the necessary information on time will have a positive effect on all the changes in the business environment. This is one of the more important opportunities that social BPM offers.

Next opportunity is addressing communication between participants. For example, monitoring email thread can be confusing at times, especially if we are not the initial recipient. Social software offers many more advanced solution for this kind of communication. Participants can use one of the group chats available, where every individual can instantly recognize the sender of the message. This opportunity of social BPM enables transparent communication, which is required for effective work. Even if individual wants to browse the previous talks, the search is usually simple and rapid.

Social BPM is in very close relation to BPM. For this reason we also highlighted opportunities, which are related to BPM. Firstly, successful BPM will require leadership to be effective, just as all other agents of change require leadership. If you are a BPM practitioner and you are willing to lead, you can make an outsized impact on your organization and your business. Secondly, social BPM offers a way for those inclined to invest in learning BPM and in improving their personal skills. Thirdly, the cloud; BPM has been (relatively) slow to move to the cloud, but the trend of moving in this direction will continue. And finally, companies need a successful partner and ecosystem strategy to make enterprise software work if it requires any integration with other enterprise systems [35].

The article [36] explains how social software allows a diverse communication and shared knowledge within a company. They used a “status feeds” to support the execution of non-predictable business processes. They developed a system to better support execution. At the end they conclude, that Web 2.0 does not consist of complex new developments but rather of a new composition of existing methods and techniques. This example can be a model for other companies as well. Technologies of Web 2.0 could therefore be components of greater and custom-made platform.

4.4 Threats

In paragraph bellow we concentrate on threats that could cause trouble in business.

When dealing with social BPM we came across very large set of actors from different organizations and with varying levels of BPM proficiency and domain knowledge [37]. Social BPM can only be successful if all actors who can make meaningful contribution are motivated to invest their time and effort continuously and lastingly. Some of participants may also have very limited skills. Software is used at different levels of BPM and is usually microblogging to broadcast status changes, or employing a Wiki for process modelling and documentation [17].

Often, the problem is in the stakeholders as well. There are typically many key stakeholders in the business, IT and operation functions in most BPM projects. Across the different BPM projects, and as the corporation embraces enterprise-wide BPM adoption, some of these stakeholders are likely to include part of the executive C-suite audience. Without clear and effective stakeholder management, even if individual BPM project succeed and deliver value, it is rare that enterprise adoption will be successful [38].

Next threat is related to security requirements. User *authorization* defines which users are allowed to operate or participate in a process. Security mechanisms should restrict permission only to required resources and data, and only for specific time period for each operating user. This requirement includes the role management of the users involved in business processes. User *authentication* assures that a process allows access only to authorized users and denies access to the rest. Successful authentication is a challenge for system designers, security experts and for the modelling tools. Business process *auditing* defines that each process is supervised in order to detect erroneous actions. Auditing involves log file analysis regarding process execution, messaging flows, communication with external processes to the environment, etc. *Confidentiality* refers to unauthorized access to systems or resources by processes or users not allowed to. *Data integrity* forms an important part of a business process. Thus, for any process that is corrupted or suddenly terminated should be guaranteed that none of the data involved are lost [39].

Another important thread to take into the account is the small number of highly active users in specific enterprise. It is important to develop a system to persuade the employees to contribute. Social software provides its benefits only after some time, therefore the employees have to be persuaded to invest time and effort for some time to achieve benefits in the future [24].

There are also some views why social BPM is not good practise [38]. First, some opponents are exposing that social BPM is the latest silver bullet and it may become just another failed IT implementation, especially if it is driven by the so called “IT geeks” and not “real” business users.

Secondly, they claim that business is inherently "social", because people interact all the time to communicate, share, learn and solve problems and to innovate. Thus, there is nothing new about social BPM.

They also exposed the rule 90-9-1, where 90 % of users read, but not contribute, 9 % users contribute from time to time, but other priorities dominate their time and 1 % of users participates a lot, by posting content. They support their claims with example where 80 % of Knowledge Management happens through human interaction and only 20 % can be captured using technology. So, the opponents doubt that IT tools will be the answer for social BPM [38].

In [6] some other threats are identified based on practical results of an experiment. Almost 50 % of all participants were reporting problems with tools. They also did not like the way they were communicating. Their chat system did not explicitly identify the members and this has bothered them.

5 Conclusion

In this article, we examined the topic of social BPM, which combines modern Web 2.0 concepts and more traditional BPM ones. Unifying both approaches enables a better collaboration and communication between participants, which improves the traditional approaches to BPM. In the article we firstly reviewed the background of both concepts, where we confirmed that both BPM and Web 2.0 are already mature and widely accepted concepts. Combining them resulted in a concept, commonly referred to as social BPM - an interesting research area, which is still in its infant stages.

The focal point of our article is a SWOT analysis, which focuses on strengths, weaknesses, opportunities and threats for a particular subject. To achieve these objectives, we reviewed the existing literature that addresses the issue of social BPM.

The results of the SWOT analysis stated that social BPM has many advantages, among such are predominant the transparency of collaboration and cooperation between various stakeholders of the project. Besides this advantages, social BPM also offers the opportunity to develop a tool for the integration of various social resources. The lack of such unification represents the greatest weakness of social BPM. Also, when implementing social BPM, the greatest risk are human factors, since people with different experiences and motivation collaborate on the same project.

The following implications can be foreseen. Businesses that are considering switching to social BPM can focus on weaknesses and threats. The results provide an insight to the downside of social BPM, so that potential adopters can prepare for a smooth transition. Also, readers that are currently discovering the aspect of social BPM have an overview of the current state of affairs regarding the topic.

It is important also to highlight that social BPM, based on discovered advantages, is an appropriate approach for those companies (or employees), who actually see potential in it. The motivation is very important; to adopt social BPM it is necessary to invest some time to learn new features.

The use of social technologies firstly has become very popular in private life, and was later mapped to business environments. The research [32] from 2012 indicates that use of social technologies in business environments has started to appear, but is represented only in 1–2 % of all employees. So, we have a reason to believe that social technologies have the potential to be adopted within companies, and adopters can gain many benefits from the use of such technologies.

Readers should consider the following limitations. Firstly, the results of SWOT analysis have been identified as difficult to translate into meaningful actions. Secondly, a different set of scientific and popular articles could provide different set of results. Thirdly, since social BPM is currently rising in popularity, the results of our SWOT analyses may quickly become outdated.

In future research we plan to conduct systematic literature review, which is focused on a research question that tries to identify, appraise and synthesize all research evidence relevant to that question. Additionally a case study of transforming traditional BPM to social BPM will be performed.

References

1. Künzle, V., Reichert, M.: Towards object-aware process management systems: issues, challenges, benefits. In: Halpin, T., Krogstie, J., Nurcan, S., Proper, E., Schmidt, R., Soffer, P., Ukor, R. (eds.) Enterprise, Business-Process and Information Systems Modeling. LNBIP, vol. 29, pp. 197–210. Springer, Heidelberg (2009)
2. Pereira, N., Vera, D., Miller, H.G.: Business Process Management and the Social Web. IT Pro, pp. 58–59, December 2011
3. Stopar, D., Bajec, M.: Upravljanje poslovnih procesov (2009)
4. Havey, M.: Essential Business Process Modelling. O'Reilly, Sebastopol (2005)
5. Hauder, M.: Bridging the gap between social software and business process management: a research agenda. In: 7th International Conference on Research Challenges in Information Science, pp. 1–6 (2013)
6. Caporale, T., Citak, M., Lehner, J., Schoknecht, A., Ullrich, M.: Social BPM lab – characterization of a collaborative approach for business process management education. In: 2013 IEEE 15th Conference Business Informatics, pp. 367–373 (2013)
7. Vaz Vieira, A.R., Jaklič, J.: Business process management and social networks : a case study in an airline organization. In: The International Scientific Conference on Management of Knowledge and Learning MakeLearn, pp. 151–158 (2013)
8. Rouse, M.: Business process management (2011). http://searchcio.techtarget.com/definition/business-process-management
9. Pflanzl, N., Vossen, G.: Human-oriented challenges of social BPM : an overview. EMISA **222**, 163–176 (2013)
10. Weske, M.: Business Process Management: Concepts, Languages, Architecture. Springer, Heidelberg (2007)
11. Rouse, M.: Web 2.0 (2011). http://whatis.techtarget.com/definition/Web-20-or-Web-2
12. Quarterly, T.M.: How businesses are using Web 2 . 0 : A McKinsey Global Survey (2007)
13. Evaluation, P., Languages, P.M., Systems, I.: Ontology - vs. pattern-based evaluation of process modeling languages: a comparison. Communications **20**, 774–799 (2007)
14. Mcafee, A.P.: Enterprise 2.0: the dawn of emergent collaboration. Manag. Technol. **47**(3), 21–28 (2006)
15. Curtis, A.: The brief history of social media (2013). http://www2.uncp.edu/home/acurtis/NewMedia/SocialMedia/SocialMediaHistory.html

16. Milgram, S.: The small world problem. Psychol. Today **1**(1), 61–67 (1967)
17. Silva, A.R., Rosemann, M.: Processpedia: an ecological environment for BPM stakeholders' collaboration. Bus. Process Manag. J. **18**(1), 20–42 (2012)
18. Schmidt, R.: A framework for the support of value co-creation by social software. In: Daniel, F., Barkaoui, K., Dustdar, S. (eds.) BPM Workshops 2011, Part I. LNBIP, vol. 99, pp. 242–252. Springer, Heidelberg (2012)
19. Fleischmann, A., Schmidt, W., Stary, C.: Subject-oriented BPM = socially executable BPM. In: 2013 IEEE 15th Conference Business Informatics, pp. 399–407, July 2013
20. Surowiecki, J.: The Wisdom of Crowds, p. 336 (2005)
21. Rangiha, M.E, Karakostas, B.: A socially driven, goal-oriented approach to business process management. Int. J. Adv. Comput. Sci. Appl. (2013)
22. Richardson, C.: Is Social BPM a Methodology, a Technology, or Just a Lot of Hype? Blog (2010). http://blogs.forrester.com/clay_richardson/10-05-20-social_bpm_methodology_technology_or_just_lot_hype
23. Schmidt, R., Nurcan, S.: BPM and social software. In: Ardagna, D., Mecella, M., Yang, J. (eds.) BPM 2008 Workshops. LNBIP, vol. 17, pp. 649–658. Springer, Heidelberg (2009)
24. Erol, S., Granitzer, M., Happ, S., Jantunen, S., Jennings, B., Johannesson, P, Koschmider, A., Nurcan, S., Rossi, D.: Combining BPM and social software : contradiction or chance? pp. 449–476, April 2010
25. Clases, C., Bachmann, R., Wehner, T.: Studying trust in virtual organizations. Int. Stud. Manag. Organ. **33**(3), 7–27 (2003)
26. Sydow, J.: Dynamik von Netzwerkorganisationen–Entwicklung, Evolution, Strukturation, Organ. Theor. und Praxis. Stuttgart, pp. 327–356 (2003)
27. Back, A., von Krogh, G., Seufert, A., Enkel, E.: Putting Knowledge Networks into Action. Springer, Heidelberg (2005)
28. Swenson, K.: The quantum organization: how social technology will displace the newtonian view. In: Social BPM: Work, Planning and Collaboration Under the Impact of Social Technology, p. 234 (2011)
29. Filipowska, A., Kaczmarek, A., Koschmider, M., Stein, S., Wecel, K., Abramowicz, W.: Social software and semantics for business process management - alternative or synergy? J. Syst. Integr. **2**(3), 54–69 (2011)
30. Earls, A.: Social BPM, collaborative BPM: sorting out which to use when (2012). http://www.ebizq.net/topics/social_bpm/features/13347.html. Accessed 29 January 2015
31. Kemsley, S.: Business process management meets enterprise 2.0 (2010). http://www.slideshare.net/skemsley/business-process-management-meets-enterprise-2-0
32. McConnell, M.: The pros and cons of social business (2013). http://www.itbriefcase.net/it-briefcase-interview-the-pros-and-cons-of-social-business. Accessed 29 January 2015
33. Swenson, K.: Who is socializing in social BPM? (2010). http://social-biz.org/2010/05/12/who-is-socializing-in-social-bpm-2/
34. Francis, S.: Discussions of bpm opportunities for 2015 (2014). http://www.bp-3.com/blogs/2014/12/discussions-of-bpm-opportunities-for-2015/
35. Pflanzl, N., Vossen, G.: Challenges of social business process management. In: 2014 47th Hawaii International Conference System Sciences, pp. 3868–3877 (2014)
36. Aggarwal, S.: Common pitfalls of an enterprise BPM program. http://www-304.ibm.com/connections/blogs/aim/entry/common_pitfalls_of_an_enterprise_bpm_program?lang=sl_si
37. Gritzalis, D., Stavrou, V., Kandias, M., Stergiopoulos, G.: Insider threat: enhancing BPM through social media. In: 2014 6th International Conference on New Technologies, Mobility and Security (NTMS), pp. 1–6 (2014)

38. Seath, I.J.: Social BPM: more questions than answers? (2011). http://ianjseath.wordpress.com/2011/10/03/social-bpm-more-questions-than-answers/
39. Graupner, S., Bartolini, C., Motahari, H., Mirylenka, D.: When social media meet the enterprise. In: Proceedings of the 2012 IEEE 16th International Enterprise Distributed Object Computing Conference EDOC 2012, pp. 201–210 (2012)

Understanding a Company's Social Media Strategies and Customer Engagement

Eric K.W. Lau(✉)

Department of Marketing, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong
mgeric@cityu.edu.hk

Abstract. Social media is becoming imperative in today's Internet world. A real case study of a company's social media strategies was presented, so as to understand the importance of social media in marketing management. The present study, therefore, is an exploratory attempt to analyze the company's social media strategies and their corresponding marketing impact. The collaborative processes of action research, where content analysis explains reflections on the lived marketing experience in Facebook, develops knowledge of complicated consumer behavior, such as customer engagement in social media.

Keywords: Social media · Customer engagements · Interpersonal bonds

1 Introduction

Marketing has undergone dramatic changes in the last decades. For example, social media has changed the marketing functions across many industries in recent years. Marketers can now connect with customers online, record every click on the web, watch every step in a store, listen to all public conversations, and understand their preferences before marketing campaigns. It is not difficult for marketers to know what consumers think, feel, and do in the consumption process from social media [13]. In Taiwan, for example, social media marketing is getting more and more popular. However, relatively little research has been directly devoted to developing contemporary approaches so as to build up the relationship between a company's social media activities and customer engagement. The present study is an exploratory attempt to analyze the influence of social media on marketing management, using a real company case. The investigation may help marketers to manage customer engagement more effectively in its social media strategies.

2 Popularity of Social Media

In the 1990s, when the Internet began to be commonly used by the general public, it was mainly in the form of e-mail communication. Afterwards, as broadband Internet connection began to be widely used, companies began to use websites to communicate with their customers. The Internet has provided many benefits to consumers. The most well-known benefits are that the Internet speeds up the communication process,

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 438–450, 2015.
DOI: 10.1007/978-3-319-21009-4_34

including consumer-to-consumer information flow. Online communities are the product of Internet applications [22].

Social media is now becoming imperative in today's Internet world. It is defined as "a group of internet-based applications that build on the ideological and technical foundations of Web 2.0, and that allow the creation and exchange of user generated content" [19, p. 61]. It allows individuals to create, collaborate, and share online content with other persons. There are lots of well-known social media platforms commonly used, such as Facebook, Youtube, Instagram, Twitter, LinkedIn, Weibo, Wechat, and others. In 2013, the number of Facebook users, for example, reached 1.11 billion within nine years of the platform being launched [10]. Messages discussed in the social media platforms are mostly personal, but some are commercial posts. Originally, those messages were referred to as an individual's online diary, which may contain personal opinions toward specified topics or factual information. Most of the personal messages posted in social media platforms are complementary, and visitors are allowed to comment on their posts in real time. It tends to be transparent, and it is available to everyone on the Internet to view, comment on, and share. Messages in the social media platform are commonly "follow," "like," and "share" among people [24]. Social media brings all people on the Internet together to act collectively. It opens up personal opinions, ideas, and feelings to other people in the free-form writing space online. Consumers map their knowledge from social media and peer-to-peer reliance.

3 The Importance of Social Media Marketing

In the views of many people, shopping is an important social and personal activity [14]. Attributes of the shopping experience are important determinants of patrons' behavior. Another aspect of participation in social media is compatibility with consumers' lifestyles and shopping habits—a main factor associated with the sense of belonging. It is proposed that while assessing the compatibility of an innovation, one looks at its compatibility with existing values and beliefs, previously introduced ideas, and potential consumer needs. Bellman and colleagues [4] argue that people living a wired lifestyle patronize online stores spontaneously. These consumers use the Internet as a routine tool in their daily and work life. Their habitual use of the Internet for other purposes leads them to naturally use it as an information channel as well. It was suggested that customers tend to expect companies they trust to interact with them in social media [9].

Company reputation, therefore, is heavily linked with consumer buying processes. It consists of information related to the behavior, history, values, prestige, and characteristics of the brand involved in the transaction. Customers map their knowledge from media reports and peer-to-peer reliance. They will then assess the likelihood of the particular brand involved in the buying decision. These answers are the fundamental rule of how hard the brand will need to work to build customer trust and loyalty. People change their attitudes, beliefs, and even behaviors in reaction to real or imagined social influence. It was suggested that personal behavior is learned from peers or near-peers [3]. Normative obligation is the inevitable end result of institution [23]. It is just a social pressure places on individuals. "Normatively appropriate" and "taken-for-granted structure" from institutions can be two guiding principles for individuals making a purchase decision. People's perceptions of a particular product or service, therefore, are partially shaped by the existing value system of a society. If people's perceptions are inconsistent with others, there is a greater tendency for people to change their attitudes as to match others [6, 12]. If everyone else is behaving in a certain way, then observers conclude that it must be appropriate. Buyer communities have traditionally operated through social interaction. Consumers hold to ideal desires and expectations in the buying process. The reputation of those judged to be liable will vanish, and these consequences are far greater than the gain they have reaped. "Good" and "bad" are the social evaluations, rather than objectively defined cognitive objects. Interestingly, this suggests that the benefits or pain of buying something may not come from knowing the person actually buying it, but rather from their social interactions with social media. Therefore, social media can be full of marketing knowledge (i.e., customer preference) for marketers [27].

Particularly, social media enhances customer engagement [28, 30]. It forms the new construct, i.e., interpersonal bonds between company and among different customers. Companies that focus on the interpersonal relationship aspects of the exchange can keep customers in the relationship. The development of interpersonal bonds may be a way for a company to differentiate itself from other competitors.

4 Hypothesized Themes of the Company's Posts in Social Media

Attitude is considered one of the most important constructs in consumer behavior. It is generally defined as a learned disposition to respond in a consistently favorable or unfavorable manner with respect to a given object, issue, or behavior [11]. There are a large number of studies dedicated to identifying the variables that affect online shopping attitudes and intentions. In order to consider the relevant set of individual beliefs, the theoretical framework proposed by Jarvenpaa and Todd [15] will be adopted to identify the salient factors that affect the adoption of online shopping. The authors have identified eleven factors conditioning the adoption of online shopping by users. These factors are classified into three different categories: product perceptions, shopping experience, and customer services, with a sub-category in each. They are proposed to be three main functions commonly provided in the social media platform between a company and its customers, i.e., product perception by the company's post in social media (i.e., the context of post), shopping experience sharing among people themselves (i.e., talking about), and customer services provided by the company in social media.

Product perception serves as the primary determinant of where consumers choose to shop [8]. Among these the most significant factors appear to be perceptions of price, product quality, and product variety. They are hypothesized as the main context of a post created by the company in social media.

The price of online shopping over the Internet falls into two parts. The first part is the set-up outlay, which includes computer hardware, software, subscription fee, maintenance cost, etc. The second part will be the retail price of the goods as offered in virtual and traditional stores. In this survey "price" will be confined to the second part of the price of online shopping over the Internet. Price is the total monetary cost to the consumer of purchasing a good or service. Low pricing is viewed as particularly vital in enticing consumers to try new products or new ways of shopping [16]. Lodorfos et al. [21] support that financial incentives were remarkably the motivation for early online shoppers expecting to find the lowest prices from online merchants as a reward for their risk taking of no physical examination before purchase. Further, the belief that price is a main purchasing determinant for online buyers is reinforced by the success of "auction sites" [20]. Therefore, it is expected that a company can post a discount message on social media to stimulate a sale.

Perception of product quality refers to the extent to which the consumer believes that the web vendor provides the quality of products and services he expected [15]. It relates to those distinguishing characteristics or traits inherent in the product or service that differentiate it from competitive products or service offerings [8]. Brands and retailers that are famous and well reputed from traditional channels may translate to quality in an online channel. The question becomes how consumers will assess product quality when they are unfamiliar with the retailer or the product brand [16]. Jasper and Lan [17] emphasize the need to provide independent evaluations of goods and services to convince consumers of the product quality on the Internet. Nevertheless, product quality is always a significant aspect of the purchasing decision: its importance is intensified when purchasing over the Internet [5]. Product quality may be another good context of the message for marketers in their social media strategies.

Product variety is the assortment or range of goods available from a retailer. Merchants who offer large selections seem to be more successful [16]. There are several reasons why online shoppers value product variety. First, larger assortments can increase the possibility of the fulfillment of online shoppers' needs, especially when the product is very likely to be unavailable from conventional retail channels [29]. Second, a wider assortment with a sophisticated search engine would facilitate consumers to purchase better quality items [2]. Finally, wider product choice available online means richer information, which would lead to more sensible purchasing decisions and higher satisfaction levels [25]. It is also worth noting that lots of marketing literature explores the relationship between perceived variety and actual assortment. Most researchers consent that consumers generally prefer more variety when given a choice [18]. Further, anecdotal evidence suggests that consumers place a high value on increased product variety. For example, Yahoo/AC Nielsen's Internet Confidence Index lists "wide selection of products" as one of the top three drivers of consumer e-commerce based on a survey of Internet purchasers.

5 The Case Study

The selected company (i.e., Brand X) in the study is an international fashion brand with business operations in Taiwan. The main business of Brand X in Taiwan is selling fashion and accessories that target youngsters. Social media, particularly Facebook, was selected as the major data source for the study. The data collection period is 1 January 2014 to 31 December 2014. Casual observations of Brand X's social media activities from its Facebook page, such as number of posts per month, the number of likes, number of talking about (i.e., about people engaged and interacting with particular company posts on the Facebook page), number of influential fans, and number of shares were reviewed.

In developing the conceptual framework for the study, a preliminary qualitative study was carried out. The study involved a content analysis of 103 messages posted by the company on its Facebook page from 1 January 2014 to 31 December 2014. The posts were transcribed, coded, and interpreted in light of a priori themes drawn from the literature.

6 Basic Observations and Findings

The main objective of the study was to examine systematically the messages posted by the company on its Facebook page and to investigate customer engagement accordingly. Therefore, the study first examined the context of the messages posted by the company. The content of the verbatim transcripts of the posts was then analyzed (Table 1). The contexts of the posts that appeared frequently in the messages were identified, and these contexts were identified as content analysis variables and then sorted according to common themes. Analysis of the sample revealed 11 recurring themes in the 103 messages including the average numbers of likes, talking about, influential fans, and shares. These are listed in Table 2.

Marketing events and new product launches were found to be the dominant contexts of the posts by the company on its Facebook page (more than 26 % of the total posts created within a year). Posts regarding the company's recent marketing events, including its marketing campaigns and promotions, are commonly found on its Facebook page. Surprisingly, it is inconsistent with previous research [7]: product price or discount message is not the top priority context selected by the company for its Facebook page in the study.

The following post is a typical example that the company shared at its recent marketing event and gained 136 likes and one share on 26 December 2014 (Fig. 1). The company titled the marketing event with a short title sentence as to draw people's awareness, promoted the event briefly by using less than 100 words, and invited people to participate in the event by clicking the corresponding links. The company also posted the event photo immediately after the description. Within a week 136 fans liked the post.

Table 1. The transcripts of the posts example.

Reference	Month	Date	Number of likes	Number of talking about	Number of influential fans	Number of shares	Context category
1	1	1/3/2014	115	0	0	0	1
2	1	1/6/2014	85	2	1	0	6
3	1	1/7/2014	160	7	6	5	6
4	1	1/7/2014	68	0	0	0	6
5	1	1/9/2014	57	57	57	0	6

Table 2. Summary of the posts investigated.

Context of the post	Actual number of posts	Percentage	Average number of likes	Average number of talking about	Average number of influential fans	Average number of shares
Marketing events	27	26.21 %	175.44	5.63	6.3	7.59
New products launch	27	26.21 %	150.26	1.3	0.85	3.59
Game related	9	8.74 %	335.89	167.56	167	160.11
Marketing events (Behind the scene)	9	8.74 %	194	4	3.67	7.78
Change of facebook profile or cover photo	8	7.77 %	139.25	1.5	1.63	0.5
Greetings	6	5.83 %	117.17	0	0	0.17
Price or discounts	5	4.85 %	145	2.2	1.6	8.2
TVC/promotion video share	5	4.85 %	138.4	1.8	1.4	8
Share related magazine/news contents	3	2.91 %	147.67	0.67	0.67	0
Created in Taiwan concept	2	1.94 %	214	2	2.5	7
New store opened	2	1.94 %	128.5	4	2	0
Total	103					

Another post by the company, shown below, is about its new product launched on 18 December 2014 (Fig. 2). One hundred thirty-seven people liked the post, one person shared it on his/her Facebook page, and six people were talking about it as the comments left on the company's Facebook page. Interestingly, the company responded to the comments left by the influential fans and appreciated their support. It is a typical example of how marketers can enhance their interactions with customers in social media.

It was noted that although the company only has nine posts for the online competitions and games on its Facebook page, it created effective effects on customer engagement (average number of likes = 335.89 per post, average number of talking about = 167.56 per post, and average number of shares = 160.11 per post; both are largest numbers identified in the study). The effects were identified by the number of likes, the number of talking about, and the number of shares on the Facebook page. As shown in Fig. 3, the company posted a message asking people to vote for their preference of two promotion styles on 25 March 2014. Eventually it created an active customer engagement (1015 likes and 618 shares within a week).

Fig. 1. Marketing event post example

Fig. 2. New product post example

Another common context posted by the company on its Facebook page is some behind the scenes description of marketing events. It created a relatively large number of likes (average number of like = 194 per post, second largest number of likes identified in the study). As illustrated in Fig. 4, the company posted about some behind

Fig. 3. Game or competition post example

Fig. 4. Marketing event (behind the scenes) post example

the scenes of its marketing event on 26 September 2014. Eventually it created an active customer engagement (410 likes, 8 talking about, and 22 shares within a week).

Interestingly, the study found that people were aware that the company changed its Facebook profile picture or cover photo. In 2014, the company had eight posts about its new profile picture or cover photo, which created 139.25 likes, 1.5 talking about comments, and 0.5 shares on average. As shown in Fig. 5, the company changed its profile picture on its Facebook page and created 201 likes within a week.

In the study, it was found that a greetings post is the least useful way to increase customer engagement in social media. A relatively small number of likes was generated by the greetings (average number of likes = 117.17 per post, smallest number of likes identified in the study). As shown in Fig. 6, the company alerted people to the rainy day in Taiwan in a post with a rainy day photo on 13 August 2014, but only 115 people liked the post at the end (no talking about and no shares).

One-way analysis variance (i.e., ANOVA) was conducted to test the differences between the means of customer engagement variables (i.e., number of likes, number of

Fig. 5. Change of Facebook page profile or cover photo example

Fig. 6. Greetings post example

shares, and number of talking about) from different contexts of posts by the company (Table 3). It was found that the different context of the post generates a different number of talking about (F(10, 92) = 2.836, $p < 0.05$) and number of shares (F(10, 92) = 2.361 $p < 0.05$). However, no differences were identified about the 11 contexts of posts associated with number of likes (F(10, 92) = 1.146, $p > 0.05$). It is suggested that the selection of context of messages posted by the company can affect the number of talking about and number of shares on its Facebook page, but not the number of likes.

As shown in Table 3, it was found that the average number of likes has a positive and relatively strong association with both the average number of talking about (r = 0.693, $p < 0.001$), average number of influential fans (r = 0.700, $p < 0.000$), and average number of shares (r = 0.750, p < 0.001) on the selected company's Facebook page.

Importantly, the average number of talking about has a positive and very strong association with the average number of shares (r = 0.988, $p < 0.001$), and the average number of influential fans has a positive and very strong association with the average number of shares (r = 0.988, $p < 0.001$).

It is suggested that "like", "share," and "comment" are very important functions of a company's Facebook page to activate customer engagement.

A correlation analysis was conducted on all variables to explore the relationship between variables related to customer engagement. The bivariate correlation procedure was subject to a one-tailed test of statistical significance at two difference levels: highly significant ($p < 0.001$) and significant (p < 0.05). The Pearson correlations are shown in Table 4. The Pearson results are reported below.

Table 3. One-way ANOVA.

Construct		Sum of squares	df	Mean square	F	Significance
Number of likes	Between groups	303749.472	10	30374.947	1.146	0.337
	Within groups	2438209.441	92	26502.277		
	Total	2741958.913	102			
Number of talking about	Between groups	223306.748	10	22330.675	2.836	0.004
	Within groups	724305.970	92	7872.891		
	Total	947612.718	102			
Number of shares	Between groups	198396.089	10	19839.609	2.361	0.016
	Within groups	773017.115	92	8402.360		
	Total	971413.204	102			

Table 4. Pearson correlation analysis (significance).

Construct	1	2	3	4
Average number of likes	1.000			
Average number of talking about	0.693**	1.000		
	(0.000)			
Average number of influential fans	0.700**	1.000**	1.000	
	(0.000)	(0.000)		
Average number of shares	0.750**	0.988**	0.988**	1.000
	(0.000)	(0.000)	(0.000)	

Note: ** Correlation is significant at 0.01 level (1-tailed)

7 Managerial Implications

The communication of information about the brand should be transparent, although one should not underestimate the fact that managers must act wisely in social media marketing. Credibility is at stake. The consequences of secrecy or deception are massive: even if the company is able to survive, it is highly unlikely that it will regain the trust of its staff, public, or shareholders. In addition, social media, such as Facebook, is an important marketing channel that marketers cannot afford to ignore for customer engagement.

The study identified that a marketing event post is an important social media strategy. Customer experience is not an amorphous construct; it is as real as an offering as any service, good, or commodity. Commodities are fungible, goods tangible, services intangible, and experiences memorable. No two people can share the same experience, as each experience is derived from the interaction between the staged event and the individual's state of mind. Abbott [1] suggests that what people really desire is not products, but satisfying experiences. Pine and Gilmore [26] suggest experiences can be viewed in two dimensions. The first is about customer participation, both passive and active. The second describes the connection, or environmental relationship, that unites customers with the event or performance. Like goods and services, it has to meet customer need, to work, and to be deliverable. Disneyland is one of the most typical examples to illustrate the concept—selling experiences spreads beyond theaters and theme parks. It uses services as a stage, and goods as props, to engage individual customers in a way that creates a memorable event. It is expected that marketers can differentiate the customer experience in their social media strategies.

8 Conclusion

Social media is a hotly debated marketing issue. The demonstration in this paper provides evidence to support the concept. The study focuses on marketing management from participatory action research in social media. It used a collaborative inquiry approach. In terms of social interaction, traditional collaboration with consumers and marketers is often cited as being an important feature of customer relationship management. However, with the help of technology, interaction and collaboration are made possible even without personal contact. Indeed, this interaction and collaboration is, in many ways, more flexible and extensive than traditional marketing management.

References

1. Abbott, L.: Quality and Competition. Columbia University Press, New York (1995)
2. Bakos, Y.: Reducing buyer search costs: implications for electronic marketplaces. Manage. Sci. **43**(12), 1676–1692 (1997)
3. Bandura, A.: Social Learning Theory. Prentice-Hall, Englewood Cliffs (1977)
4. Bellman, S., Lohse, G., Johnson, E.: Predictors of online buying behavior. Commun. ACM **42**(12), 32–38 (1999)
5. Boyer, K.K., Hult, G.T.M.: Customer behavioral intentions for online purchases: an examination of fulfillment method and customer experience level. J. Oper. Manage. **24**(2), 124–147 (2006)
6. Buehler, R., Griffin, D.: Change-of-meaning effects in conformity and dissent: observing construal processes over time. J. Pers. Soc. Psychol. **67**(6), 984–996 (1994)
7. Chen, Y., Fay, S., Wang, Q.: The role of marketing in social media: how online consumer reviews evolve. J. Interact. Mark. **25**(2), 85–94 (2011)
8. Cronin, M.J.: Global Advantage on the World Wide Web. Van Nostrand Reinhold, New York (1996)
9. Dutot, V.: A new strategy for customer engagement: how do french firms use social CRM? Int. Bus. Res. **6**(9), 54–67 (2013)
10. Facebook (2013). www.facebook.com
11. Fishbein, M., Ajzen, I.: Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research. Addison-Wesley, Massachusetts (1975)
12. Griffin, D.W., Buehler, R.: Role of construal process in conformity and dissent. J. Pers. Soc. Psychol. **65**, 657–669 (1993)
13. Hennig-Thurau, T., Hofacker, C.F., Bloching, B.: Marketing the pinball way: understanding how social media change the generation of value for consumers and companies. J. Interact. Mark. **27**(4), 237–241 (2013)
14. Holt, D.B.: How consumers consume: a typology of consumption practices. J. Consum. Res. **22**(1), 1–16 (1995)
15. Jarvenpaa, S.L., Tractinsky, N., Saarinen, L.: Consumer trust in an Internet store: a cross-cultural validation. J. Comput. Mediated Commun. **5**(2), 1–35 (1999)
16. Jarvenpaa, S.L., Todd, P.A.: Is there a future for retailing on the Internet. In: Peterson, R.A. (ed.) Electronic Marketing and the Consumer, pp. 139–156. Sage Publications, Thousand Oaks (1997)
17. Jasper, C.R., Lan, R.R.: Apparel catalog patronage: demographic, lifestyle and motivational factors. Psychol. Mark. **9**(4), 275–296 (1992)
18. Kahn, B.E., Lehmann, D.R.: Modeling choice among assortments. J. Retail. **67**(3), 274–299 (1991)
19. Kaplan, A.M., Haenlein, M.: Users of the world, unite! the challenges and opportunities of social media. Bus. Horiz. **53**(1), 59–68 (2010)
20. Kung, M., Monroe, K.B., Cox, J.L.: Pricing on the Internet. J. Prod. Brand Manage. **11**(4/5), 274–287 (2002)
21. Lodorfos, G.N., Mulvana, K.L., Temperley, J.: Consumer behaviour: experience, price, trust and subjective norms in the OTC pharmaceutical market. Innovative Mark. **3**, 41–66 (2006)

22. Luppicini, R.: Categories of virtual learning communities for educational design. Q. Rev. Distance Educ. **4**(4), 409–416 (2003)
23. Meyer, J., Rowan, B.: Institutionalized organizations: formal structure as myth and ceremony. In: Powell, W., DiMaggio, P. (eds.) The New Institutionalism in Organizational Analysis, pp. 41–62. The University of Chicago Press, Chicago (1991)
24. Mitic, M., Kapoulas, A.: Understanding the role of social media in bank marketing. Mark. Intell. Plan. **30**(7), 668–686 (2012)
25. Peterson, R.A., Balasubramanian, S., Bronnenberg, B.J.: Exploring the implications of the internet for consumer marketing. J. Acad. Mark. Sci. **25**(4), 329–346 (1997)
26. Pine, B.J., Gilmore, J.H.: Welcome to the experience economy. Harvard Bus. Rev. **76**, 97–105 (1998)
27. Sarabdeen, J.: Legal risks in social media marketing. Int. J. e-Educ. e-Bus. e-Manage. e-Learn. **4**(3), 218–223 (2014)
28. Sashi, C.M.: Customer engagement, buyer-seller relationships, and social media. Manage. Decis. **50**(2), 253–272 (2012)
29. Sin, L., Tse, A.: Profiling internet shoppers in Hong Kong: demographic, psychographic, attitudinal and experiential factors. J. Int. Consum. Mark. **15**(1), 7–29 (2002)
30. Wirtz, B.W., Schilke, O., Ullrich, S.: Strategic development of business models: implications of the web 2.0 for creating value on the internet. Long Range Plan. **43**(2–3), 272–290 (2010)

Using of Text Mining in Online Public Access Catalog

Zuzana Juchová Jurošková and Anna Závodská(✉)

University Science Park, University of Žilina, Univerzitná 1, 010 26 Žilina, Slovak Republic
{zuzana.juchova,anna.zavodska}@uvp.uniza.sk

Abstract. One of the products of the information society is an excessive amount of information generated by mankind, and consequently the ways to search, use, process, present and store this information. We can consider entering a keyword to search engine/search tool as a most common way of searching for information. The user consequently expects simple decision-making process when selecting an object from a found set of records. The main factor influencing the popularity and reusability of particular tool or service is interpretation of content of search results as well as form of their presentation. Libraries are still key sources of information as they make the database objects available (printed and electronic) through website and mobile applications. To search for objects they use online public access catalog (OPAC) that allows users to search and order objects from any computer or mobile device connected to the Internet. This study focuses on obtaining the most desirable bibliographic elements of users and subsequently on identifying actually provided number of keywords in the Slovak library. As this number was completely insufficient in decision-making process when selecting an object by user we design and test the possibility of using text mining to obtain keywords appropriate for filling the database in online public access catalog. Consequently, we compare these generated keywords with the real metadata elements in the Slovak Library database and identify the percentage of their conformity. Based on the results of the study we recommend creating algorithm by the specialists in the field of library and information science. This algorithm will assign a weight to individual words, for example according to their location within a document (title of the publication, chapter title, abstract, document text) and thus create keywords (the meta-data elements) in the database of online public access catalog.

Keywords: Text mining · Online public access catalog · Slovak library · Library and information research · Keywords · Elements

1 Introduction

An online public access catalog (OPAC) is a database composed of bibliographic records describing books and other materials owned by a library or a library system, accessible via public terminals or workstations usually concentrated near a reference desk to make it easy for users to request the assistance of a trained reference librarian. Most online catalogs are searchable by author, title, subject, and keywords and allow users to print, download, or export records to an e-mail account [9].

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 451–462, 2015.
DOI: 10.1007/978-3-319-21009-4_35

Currently, a large number of librarian databases contain a lot of elements of different meta-data schema describing physical objects. With the mass digitization the number of elements is still growing because of an increasing number of databases containing elements describing digital objects. For the user, it is difficult to become familiar with the amount of information provided by various databases.

People can use a portal to search for catalogs and collections that are owned by every library in Slovakia. The name of this portal is **Slovak Library**. This portal uses integrated library-information system called Virtua on whose object description was chosen MARC 21 format for bibliographic data [6].

MARC record consists of three elements [7]:

- "the record structure is an implementation of the international standard Format for Information Exchange (ISO 2709) and its American counterpart,
- the content designation (the codes and conventions established explicitly to identify and further characterize the data elements within a record and to support the manipulation of that data) is defined by each of the MARC formats,
- and the data content of the record. The content of the data elements that comprise a MARC record is usually defined by standards outside the formats. Examples are the International Standard Bibliographic Description (ISBD), Anglo-American Cataloguing Rules, Library of Congress Subject Headings (LCSH), or other cataloging rules, subject thesauri, and classification schedules used by the organization that creates a record. The content of certain coded data elements is defined in the MARC formats (e.g., the Leader, field 007, field 008)."

"The MARC 21 Format for Bibliographic Data: Including Guidelines for Content Designation defines the codes and conventions (tags, indicators, subfield codes, and coded value that identify the data elements in MARC bibliographic records. MARC 21 Format for Bibliographic Data is designed to be a carrier for bibliographic information about printed and manuscript textual materials, computer files, maps, music, continuing resources, visual materials, and mixed materials. Bibliographic data commonly includes titles, names, subjects, notes, publication data, and information about the physical description of an item" [7].

Since 2012 the Slovak National Library has been going through the **mass digitization of objects**. In the process of digitization each digital object (i.e. a digital copy of the object placed in the Slovak Library) will contain at least two elements: persistent identifier and metadata Dublin Core, however in the project "Digital Library" the format Dublin Core and MARC 21 (which will be transformed through MODS to digital repository while maintaining all descriptive metadata) have to be used.

An object in a digital library of text documents and library collections is a separate physical unit of library's collection which, in general, can have these characteristics: incremental number, physical/hard copy of the book, newspaper edition, number of printed newspaper, number of printed magazine, paper from printed scientific journal or monograph, from printed scientific proceedings, printed card as a unit of collection, photography as a unit of collection or album, printed map as a unit of collection, printed music as a unit of collection, printed graphic as a unit of collection or computer software [3].

The most widely used and internationally accepted metadata scheme Dublin Core contains the following elements: title, subject, description, type, source, relation, date, coverage, creator, publisher, relationship, contributor, rights, date, format, identifier, language, audience, provenance, etc. [1].

Currently, there has been a great deal of discussion about the relationship between the bibliographic elements and metadata elements. We see the essential difference in use - metadata elements are primarily used for searching and obtaining electronic resources. For that reason there is an effort to create metadata standards as simple and understandable representation of the elements. Metadata elements are part of digital library system, which are primarily intended for the user but also serve to describe web pages, documents, etc. in order to search them easier.

Thus it was necessary to determine which bibliographic elements are the most important from a user perspective and should be maintained for digital objects. Metadata elements are crucial for the user in the process of deciding on the choice of the object from the database. Based on the available metadata elements, however, the user often cannot decide which object is relevant to him.

It was therefore necessary to propose the possibilities of obtaining value of bibliographic elements in the digitization of objects in order to maximize the quality and quantity of provided meta-data elements value of the physical objects by the database of the Slovak Library from a user's perspective.

In the study, we did not address the fact how found objects should display on the online library catalog but what metadata scheme should contain and how to provide user with the maximum amount of filled value of metadata elements that facilitate the decision-making process when selecting an object from a searched set of results.

For the purposes of further conducting of our survey we specified the perception of bibliographic element and metadata element as follows: a **bibliographic element** with recorded value, which is available in the virtual environment of the Slovak Library database and is intended for search, becomes **metadata element** with the new ability to provide descriptive information about cultural heritage to the user. At the same time the term **most used elements** refers to all metadata elements identified within the qualitative research that the user uses in process of selecting a specific object from the database of the Slovak Library.

2 Research Approach and Methodology

We used a combination of qualitative and quantitative methods for data collection. The main objective of this study is to use advantages of both methods. Research was conducted by one author as a part of her dissertation thesis [2]. In the pre-research phase we used analysis, synthesis, and questionnaire method by which we defined the most desirable bibliographic elements by the user from which we then selected metadata elements provided by the Slovak Library. In this study we used:

– group discussion in which we determined the order of particular bibliographic elements according to their importance in the process of choosing of searched term people enter into search engine,

– data mining for identification of frequency of metadata elements most desirable by users that are provided by online databases of the Slovak Library,
– experiment of suitability of the proposed software based on comparison of results obtained by text mining with the results from the database of the Slovak Library.

3 Conducting Studies

Prior to the realization of our study we conducted pre-research, in which we obtained the necessary input data. Subsequently the study consisted of three phases, which are described in detail in the following section.

Pre-research. Within the pre-research we used analysis of metadata schemas Dublin Core, Metadata Object Description Schema (MODS), Encoded Archival Description (EAD), Europeana Semantic elements (ESE) and Machine-Readable Cataloging Standards (MARC 21) and subsequent synthesis of results needed to identify the 46 metadata elements used to describe text document in databases. We chose 90 respondents (bachelor students of mediamatics and cultural heritage in 2012 studying at the Faculty of Humanities of the University of Žilina) to fill in the questionnaire in the pre-research phase. Through the questionnaire method we identified bibliographic elements that are critical for user in the process of selecting a specific object from the library database. Through analysis we defined set of the most desirable bibliographic elements that matter to user. Subsequently we selected particular metadata elements from this set that match the ones provided by the Slovak Library.

In the process of selecting a specific object from database users mostly requested the following 10 alphabetically ordered bibliographic elements used to describe the physical text document: ABSTRACT, ISSUE DATE, MAIN TITLE, ISBN/ISSN, LANGUAGE, KEYWORDS, AUTHOR'S NAME, PUBLISHER, CONTENTS (chapter name), and SUBTITLE.

Since all users' preferred bibliographic elements are consistent with the generally recommended requirements for the description of the object processed in MARC 21 format they could be used in the next stage.

First Phase of Study. The aim of the first phase of the study was to organize bibliographic elements according to their importance in the actual process of choosing the object by user (these elements form a group of the most desirable bibliographic elements) and identify the available bibliographic elements provided within each type of record about object in the Slovak Library that users actually used in decision-making process when selecting an object. At this stage, we used a method of group discussion with open questions which was recorded by voice recorder and subsequently processed by a written transcript of the responses in software Audio Notetaker 2.5. The duration of the group discussions did not exceed the maximum recommended time - 120 min. Research sample consisted of 10 bachelor students of the University of Zilina attending internal form of studies in the study program Mediamatics and cultural heritage. Their average age was 21 years. 60 % of the research participants created women.

Research sample was intentionally selected in the pre-research phase as well as for the group discussion. We chose students of library and information studies because they acquired basic knowledge in the field of operating a memory institution as well as skills in the use of the latest information and communication technologies in educational process. However, they did not attend any professional education directly focused on the examined issue so they can be considered as sufficiently informationally and medially literate but not experts in understanding and evaluating mechanism of creating metadata elements in database systems. Part of their studies is the need to use relevant sources in elaborating their seminar papers so they come into contact with library's catalogs. Given the above characteristics we can consider chosen students as an interface between professional and non-professional users and therefore representative sample for determining user expectations on frequency of the metadata elements.

Within the questions we focused on the detection of whether respondents use Slovak Library portal [8] when searching for information. All respondents answered yes to our question, even the majority of respondents attended lectures about an expert search in online database of the Slovak Library. Within the group discussions we revealed the following facts:

- AUTHOR and TITLE of publication are crucial in the process of choosing a particular object of bibliographic elements and are considered as the main assessment criteria of relevance for users within the bibliographic elements of shortened object record of the Slovak Library (AUTHOR, TITLE, PART/VOLUME, PUBLISHER, SCOPE, MULTIMEDIA, EDITION, and ISBN).
- Bibliographic elements OBJECT, SUBTITLE AND CO-AUTHOR were identified as other critical elements for respondents using full record from database in the decision-making process of selecting an object (elements of shortened record supplemented by the CO-AUTHOR, SUBTITLE, LIABILITY, DESCRIPTION, OBJECT, and SIGNATURE).
- Respondents marked OBJECT as an uncertain element because it was a lot of abbreviations in this element as well as respondents did not perceive this element as keyword or a set of words that characterizes the content of a document and so facilitate their decision-making process.
- Respondents would welcome extension of a Slovak Library list for element marked as KEYWORD (or element, which would contain keyword of object), or elements CONTENT and ABSTRACT of text document, which would help them to make the selection process easier in the case that they do not know the specific value of the bibliographic element uniquely identifying one object in the Slovak Library. Offered list should be extended for elements that would be characterized as those familiar from the web environment, i.e. reviews from critics and experts, comments and reviews from reader as well as the possible indication of whether the book is also available in a digital version.
- Next, we investigated whether respondents prefer searching by entering specific metadata element value or keyword when searching for objects in the database of the Slovak Library. Our aim was to find out what method they use more and why. All respondents agreed that they use this procedure of searching: "I enter name of

author, title or ISBN (whatever I know) because I almost immediately find the desired record. If I do not know any of the above mentioned value I enter keywords and decide according to the value of bibliographic elements." Respondents as Internet users are used to search by keywords, however, it is complicated for them to use this method in the library database search because of the problematic subsequent selection of searched text documents (it contains an insufficient number of key information about the object to be able to choose a relevant document). It should be noted that the user has improved in the choice of keywords but it would be better to have information about text documents in the database enhanced by keywords.

– Respondents identified order of the individual bibliographic elements according to their importance in the process of selecting the object by searching through keyword. Within the pre-research all respondents clearly identified bibliographic element KEYWORD that would mostly help them in selecting an object from the database of all bibliographic elements which were identified as crucial in the process of selecting a particular object from library (ABSTRACT, DATE OF ISSUE, THE MAIN TITLE, ISBN/ISSN, LANGUAGE OF PUBLICATION, KEYWORDS, AUTHOR, PUBLISHER, CONTENT - chapter titles, SUBTITLE). They would also welcome if keywords were written in the SUBJECT part or in a dedicated section.

Given that the individual names of the bibliographic elements are not identical with the names of bibliographic elements provided by MARC 21 format, we have tried to find their alternatives in the selected format for the purposes of using the results of qualitative research in the next phase of our study (Table 1).

Second phase of study. At this stage we have identified number of the users' most desirable metadata elements provided by the Slovak Library online database through

Table 1. Assigning of bibliographic elements to elements in the MARC 21 format

Keywords	tag 650 - Subject added entry – Topical term
Abstract	tag 520 - Summary, etc. note, indicator 3 - Abstract
Contents	tag 505 – Formatted contents note, subfield $a Formatted contents note
Main title	tag 245 – Title statement, subfield $a Title proper
Name of the author	tag 100 – Main entry – personal name, tag 110 - Main entry – corporate name, tag 111- Main entry – meeting name, tag 130 - Main entry – uniform title
Language	tag 041 – Language code
Issue date	tag 260 – Publication, distribution, subfield $c Date of publication, distribution
ISBN	tag 020 – International standard book number (ISBN) subfield $a International standard book number (ISBN)
Publisher	tag 260 – Publication, distribution, subfield $a Place of publication, distribution and $b Name of publisher, distributor
Subtitle	tag 245 – Title statement, subfield $b Remainder of title

data mining process. It represents the number of completed metadata elements of individual objects in the database. As statistical file we selected a set of objects in the database of the Slovak Library, i.e. 4 203 474 records. Given the size of the database itself and the criteria necessary for the completion of the study, we have generated (on the basis of the following criteria) 119 858 records of text documents in the database of the Slovak Library from the ORACLE database system.

Our criteria for selection were all records that have completed the following tags:

- 040 – Cataloging source (PN)
- and at the same time 080 – Universal decimal classification number (NO)
- and at the same time 100 – Main entry – personal name (NN)
- or 110 - Main entry – corporate name (NN)
- or 111 - Main entry – meeting name (NN)
- or 130 - Main entry – uniform title (NN)
- and at the same time 245 – Title statement (PN)
- and at the same time (500 – General note (NO)
- or 520 – Summary, etc. (NO))
- and at the same time (600 – Subject added entry – personal name (NO)
- or 650 - Subject added entry – topical term (NO)
- or 653 – Index term - uncontrolled (NO))
- and at the same time 260 - Publication, distribution, etc.

In the next section we present only the results concerning the most desirable bibliographic element OBJECT. Amount of metadata element OBJECT was 849 409. Detailed analysis of the results of qualitative research showed that this number does not provide sufficient evidence for users since respondents identified that they did not expected any abbreviations or other unclear data in the most requested bibliographic element OBJECT.

According to the methodology of processing documents in MARC 21 format topical term should be completed in the field 650 Subject added entry – topical term. Recommendations for filling in this field are [4]:

- Topical subject added entries shall be governed by the principles of cataloging policy of the Slovak Library. A geographic name, or the name of a corporate body used in a phrase subject heading are also recorded in field 650.
- Second indicator refers to thesaurus in subject added entries.
- Second indicator with value 4 - Source not specified is not indicated in $2.
- Second indicator with value 7 - Source specified in subfield $2, where $2 contains “SNKPH“, represents authorities in the Slovak Library. If the term does not occur in a list of the Slovak Library authorities participants of Virtua catalog use their own location code (e.g.: SVKPOPH).
- According to the cataloging policy of the Slovak Library clarifying library facts are not used in this field.
- In descriptor’s type systems (participants of library management system Virtua) the necessary information is divided into the appropriate fields: 648 (chronological term), 650 (topical term), 651 (geographic name).

Table 2. Number of metadata elements – tag 650

Tag 650	
Number of tag 650 subfield $a without abbreviations	73 930
$x General subdivision (O)	82 436
$y Chronological subdivision (O)	2 580
$z Geographic subdivision (O)	4 297
Sum of the most desired keywords	163 243

Despite the fact that value of metadata elements are filled in the statistic sample size according to the methodology, they are unnecessary or difficult to understand for the user.

Therefore, we have filtered out all of the value located in the appropriate tag and removed users' unwanted value of metadata elements (Table 2). Number of results was then preferred by the user.

Average number of user's desired keywords is only 1,36 of keyword on one object of statistic sample size. Given this identified average number of keywords in a database filled according to user requirements, it is necessary to clearly increase the amount and thereby facilitate the user decision making.

Third phase of study = Use of text mining in library databases. Based on the results obtained in previous stages of research together with an analysis of theoretical knowledge gained by literature studying, we identified a text mining method as one of potentially useful for obtaining value of metadata elements needed to complete the Slovak Library databases, as well as their subsequent use for the needs of creation of digitized object record. Provided methods of classification, clustering and text extraction by text mining software were key criteria for selecting Rapid Miner software.

By an analysis of individual records of objects we selected a set of 21 documents from the provided library database as an object entering into text mining process. Criteria for their deliberate choice were completion of the field 650 – Topical term as well as the need of the existence and availability of digital copies of these objects. Given the fact that digitization of the Slovak Library documents is still in progress (digitization started in 2012 and the expected end should be in 2015) we were not able to increase the number of 21 documents because up to now the Slovak Library has been unable to provide more than this number of digital copies to the public. Moreover, the Slovak Library was closed in 2014 because of the digitization.

The specific objective of this choice was the comparison of results - filled value of metadata elements selected from database objects and on the other hand, selected keywords by the experiment. In the actual keyword generation we used tokenization method, i.e. the process of dividing text into basic parts - tokens.

The keywords term can be used in the following areas:

– Theses writing: as a mandatory part of abstract, most often defined as a one or multiple word terms, names of organizations, names of people, etc., expressing title or content of the work.

- Websites creation: it is essential to identify main keywords describing content of the website. In this case, keyword is term that is used on the website and should be identical with the keyword that user writes to search engine when he is looking for an information. On the website keyword should be in text document or header of document within the metadata elements. When choosing keywords we are searching for answers on the questions like: What are the objectives of website? What is the target group? What will be the content of website? To facilitate our work we can use tools that serve to identify appropriate keywords. These might include, for example Wordtracker, Etarget system or the Google AdWords keywords tools. On the other hand, we can use the statistics of Naj.sk system that are listing the most frequently used keywords through which users access the website [5].
- Searching on the Internet: in this case we can describe keyword as a word or set of words that users enter into the search engine [2].
- The Search Engine Optimization (SEO) is the process of influencing the visibility of a website in a search engine's results. Keywords are crucial in the SEO optimization because their right choice may increase the better position in search engines and thus more visitors on a website.

Within our qualitative research we found out that users of the Slovak Library perceive keywords as words that express the theme or topical content of the work as well as the words that they enter when searching for objects in the database of the Slovak Library.

4 Results

We identified keywords through software on the basis of frequency of their occurrence in the document. Within the research our aim was not to find the best solution for creating keywords and modify open source software for the needs of a particular organization but identify the method by which we could enhance existing database of the Slovak Library by generating metadata elements value needed to create a new database of digital library.

Firstly we converted 21 objects of the Slovak Library to plain text so they did not contain any formatting and there was no problem with language coding. Then we processed each object in the Rapid Miner software as a separate process. By allocation of weights to individual terms in text documents we obtained a list of terms represented by the chosen vector. Option TF-IDF allows us to create a vector based on the frequency and inverted frequency of occurrence of terms (words) in the document [10]. The software allowed us to remove so-called stop-words or modify words to the basic form. However, the above mentioned features are not available for texts written in the Slovak language therefore, after generating terms, we had to ignore the conjunctions, prepositions, etc.

Using the software we identified five terms - the most frequent words in the text document which can be seen on the Tables 3 and 4. We then compared these terms with value of metadata elements of each text document and the result was 77,14 %. This value expresses how likely generated keyword is also value tag 650 - topical term in the databases of the Slovak Library.

Table 3. Keywords identified by Rapid Miner software

Keywords Rapid Miner	
Term	Count
Collaboration	31
Knowledge	20
Management	19
Tool	19
ICT	18

Table 4. Metadata elements from the database of the Slovak Library

Metadata elements from the database of the Slovak Library			
650	0	7	\a information and communication technologies \2 snkbucl
650	0	7	\a knowledge management \2 snkbucl
650	0	7	\a collaboration (information and communication technologies) \2 snkbucl

5 Possibilities of Using the Study Results in Practice

There exist precise rules for choosing keywords for search engine optimization (SEO), which are defined by specialists. These rules also refer to the placement on a webpage. When we apply these rules we can shift our website to the top positions in search results. By proper indexing and optimization of website of the Slovak Library we could increase number of its visitors as a result of the better placement in the search results.

Based on our research we recommend developing algorithm by specialists in the field of library and information science. This algorithm should be able to calculate the relevance of each word, for example, according to their location within the document and so create keywords of library objects.

In the process of keywords formation in this field we would be able to use a model that would assign different weights to words which are in the text document title, chapter titles, in the abstract of the document, or in the text and not just identifying them by frequency of their occurrence in the document, as we did in our research. We used only software algorithm, which we do not modify and we did not do any linguistic correction of found words. On the other hand, it would be appropriate to use the thesaurus with the help of which we define words with the same meaning. The options would increase the quality of the model designed for the needs of formation keywords.

The quality of identification of keywords could improve the use of a large set of documents with assigned keywords that would provide a predefined set of keywords from which we can then choose all the keywords for new documents. In this case we are talking about the use of controlled vocabulary [11].

It is important to realize that we have tried to create keywords that are demanded by users while we used only rudimentary algorithm to find the most frequent words but nevertheless we have identified up to 77,14 % identity with the value of meta-data elements provided by the Slovak Library. We believe that the keywords would be similar in the case of changing the user group. However, students create majority of site visitors in the Slovak republic so this study was primarily focused on their needs in terms of keywords search.

In the case that we would like to add identified keywords to the database of Slovak Library, it cannot certainly be a fully automated process, but we would have to complement it by the human factor - qualified library staff doing the assessment of keyword, who would sort out only those keywords that are consistent with the methodology of creation metadata elements value.

We want experts in the field of library and information science to consider whether it would be preferable to add only the tag 650 – Topical term or its alternative in one of the schemes used in digital library, or rather define new tag, which would include users' demanded keywords.

Visualization of data lately seems to be crucial in the provision of information in the web environment. Interesting possibilities for its presentation also provides the Rapid Miner software. Through the cluster chart we can represent the frequency of occurrence of terms where the size of the cluster will determine the frequency and color of the cluster easier visibility of terms with the same number.

After completing the digitization of documents in the Slovak Library we recommend to carry out another research to add missing data and remove useless data (shortcuts, etc.) from a database of the Slovak Library. The continuation of research should be focused on the identification of user satisfaction with search results in new enhanced database. From the perspective of a single method of generating keyword we suggest visually present search results in the form of:

– Mark tags that can be used to highlight parts of text. Search engine might highlight text on a web page that match the user's search terms. We can use these tags to highlight search results in every found publication.
– Graphs where the distance between publications will be identified by lines and thickness of the lines will define keyword match between found publications.

References

1. Dublin Core Metadata Initiative Wiki: User Guide/Creating Metadata (2012). http://wiki.dublincore.org/index.php/User_Guide/Creating_Metadata#Guidelines_for_the_creation_of_format_content
2. Juchová Jurošková, Z.: The Measurement of Information about Cultural Heritage. [Ph.D. thesis]. University of Žilina, Žilina, SR, p. 102 (2012)
3. Katuščák, D., et al.: Digitálna knižnica a digitálny archív: žiadosť o nenávratný finančný príspevok. Slovenská národná knižnica, Martin, SR. 2011. 96 s (2011)
4. Majerová, J., et al.: Metodika spracovania dokumentov vo formáte MARC 21 (2007). http://www.snk.sk/?vecne-spracovanie-metodicke-materialy

5. Makulová, S.: Získajte vyššiu pozíciu vo vyhľadávačoch vhodným výberom kľúčových slov. In: Newsletter (2007). http://www.elet.sk/?seo-optimalizacia-1&sprava=ziskajte-vyssiu-poziciu-vo-vyhladavacoch-vhodnym-vyberom-klucovych-slov
6. Official website of The Library of Congress: MARC 21 Format for Bibliographic Data (2014). http://www.loc.gov/marc/bibliographic/
7. Official website of The Library of Congress: MARC 21 Format for Bibliographic Data: Including Guidelines for Content Designation (2006). http://www.loc.gov/marc/bibliographic/bdintro.html
8. Official website of The Slovak Library Catalog: Catalogs and collections of the Slovak Library (2011). https://www.kis3g.sk/
9. Reitz, M.J.: Online Dictionary for Library and Information Science: OPAC (2015). http://www.abc-clio.com/ODLIS/odlis_o.aspx
10. Robertson, S.: Understanding inverse document frequency: on theoretical arguments for IDF. J. Documentation **60**(5), 503–520 (2004). http://www.emeraldinsight.com/journals.htm?articleid=864256&show=abstract
11. Witten, I.H.: Text mining (2004). http://www.cos.ufrj.br/~rick/gc2010/_papers/aula13/04-IHW-Textmining.pdf

State of the Art Construction Based on the J48 Classifier: Case Study of Internet of Things

John P. Anzola Anzola[1], Luz Andrea Rodriguez Rojas[1(✉)], and Giovanny M. Tarazona Bermudez[2]

[1] Fundación Universitaria Los Libertadores, Cra. 16 # 63A – 68, Bogotá, Colombia
{jpanzolaa, larodriguezrl}@libertadores.edu.co
[2] Universidad Distrital Francisco José de Caldas, Cra. 7 no 40-53, Bogotá, Colombia
gtarazona@udistrital.edu.co

Abstract. The bibliometric technics allow us to illustrate aspects of interpretation of data obtained through data bases that are focus in science spreading, which offers an aggregated value to researchers. In this article, we explore the importance of such obtained information from articles' query; in this particular case study the Internet of Things (IoT). The obtained and preprocessed metadata is used to feed the decision tree classification (J48) algorythms, as a result of the training stage, we generated a decision tree/classification where the subjects of research are visualized within a four year time frame. The construction of this classification tree is based in entropy, thus, the order or disorder in which the subjects of research can be interpreted within a level of relevancy in the time frame mentioned. We conceived the visualization and interpretation as the state of the art for the subjects addressed for a particular case, which help researchers to infer tendencies within specific subjects of research.

Keywords: Bibliometry · Data mining · J48 classifier · Internet of things · State of the art

1 Introduction

Scientific and academic databases have become the backbone of a researcher, either for the development of a prototype, or to propose new fields of research. Search engines or meta-searchers are part of the basic tools that allows us to locate relevant information within databases such as IEEE Xplore, ScienceDirect, and Springer Link.

At the beginning these search engines aimed to safe keep information in the most efficient way, focusing in the study of cluster's physical structure, file security, and different kinds of indexation.

Through time, the data stored by those engines has changed from hundreds of thousands to billions, indexing nearly two millions of articles every year. The word indexation structure and data search made through lists and associations allow quickly query's response. Currently, a typical search can located around three thousand articles; such amount of information can take several weeks to be analyzed and synthetize by a

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 463–476, 2015.
DOI: 10.1007/978-3-319-21009-4_36

researcher. Such task requires the sorting and filtering of the data according to a relevancy order which, at this point, it's where the polemic challenge related to query, types of query, delivered data y results relevancy is centered.

The main objective of our research is to obtain a methodology that reduces the temporal analysis of query results of any given search subject that allows the researcher to infer tendencies from several topics dealt in a time frame. For this, we consider the tree representation that synthesizes the state of the art in any given subject.

The method and the technique used to build the state of the art is based on J48 Classification via decision tree learning, in which as a result a tree is built where leaves and branches are created in a way that allow users to infer information of development from different subjects; this the researcher is able to find leaves at greater depth within the state of the art tree which belongs to subject less approached.

This article is divided in several sections. The first deals with obtaining a data model that later will be used to make a typical search of results; the second section describes briefly the usage of classifiers we show the results for our case study; finally in the third section we show the conclusions and suggestion for future research.

2 Review of Previous Researches

Data mining and rules of association through decision tree classifiers create a field well explored where there are some applications for the data class extraction [1]. Data mining methods applied in this research are classifiers based on decision trees that have been used by [2–4].

However, this research contributes to the data classification methodology that constructs the state of the art for a research work. This methodology was applied to the topic "Internet of Things" (IoT) using as source the query results in IEEE Xplore which were used in this research.

On the other hand, in the bibliometric indexes the impact factor [5, 6] has been used as a measure of relevancy in most journals and it is calculated by the normalized expected relation between journal's size and number of citations received. Some of the mistakes we found in the use of such impact factor are:

- The normalization of different citation practices in diverse scientific fields [7].
- The biases caused for the two years temporal window.
- The lack of transparency in data used for its calculation.

Synoptic review was made of autonomous learning techniques with unbalanced data and they presented a meta-algorithms class with J48 based classifiers and F-measure as part of the processing [8].

Other researches focused on bibliometric inputs related with usage y application of the document classification in business and companies [9–13], the learning engine with unbalanced data [14], knowledge management [15–19], Algorithms theory and computational methodologies [20–25], computational electronic services [26–28], digital information management and context management for smart environments [29–34], quality based research and e-assessment [35–40], e-business and interpretation techniques [16, 41–45] to identify new concepts, theories, methods and techniques that allow its effective application on hybrid and technological emergent environments [18, 46–50]

Nonetheless, there are few researches that allow to obtain or to build a state of the art of particular subject in semiautomatic manner within a time frame. Therefore, the methodology that we proposed in this paper aims to provide a solution to this problem.

3 Data Model

In order to perform the information classification methodology in diverse scientific databases, we established a set of initial criteria. The first lies in the language of the search which will be English, therefore other results that come from databases of China, Germany, France, among others were dismissed. The second, indexation of meta-searching in scientific databases is done using concrete keywords and the searched topics and sub-topics are established in an interest keyword association environment; for this case study the terms "Internet of Things" (IoT) as root term as interest keyword was used.

One of the main activities when it comes to search on meta-searchers is the selection of keywords to search. Article indexation can be found in the databases from: excellence journals (local, national, regional and international), with or without arbitration journals, indexed journals, and magazines. This type of search is highly related with the kind of databases this research is focusing on, which leads to be Google Scholar indexation. This indexation is classified in:

According to the producer character:

- Scientific societies (American Psychological Association through PsycINFO, American Economic Association through Econlit)
- Academic produced (UCLA's HAPI, John Hopkins University's project MUSE, UNAM's Clase).
- Commercial produced (Elsevier, Wiley, Springer).
- Technological partners (PPCT, Redalyc, Dialnet).
- State related: National, regional and international (SciElo.org, eRevistas, ERIH)

According to content:

- With bibliometric analysis (WoS, SCOPUS, Scielo.org, SJR).
- Full text and strict requirements for articles to be published (Elsevier, Sage, Sprin-ger, Taylor & Francis, Willey, JSTOR, Redalyc, SciELO.org).
- Bibliographic or referential and abstracts (Philosofer's Index, Sociological Abstracts, Historical Abstracts, GEOBASE, Econlit, etc.).
- Directories (Ulrich's, Latindex, DOAJ).

The search of different modules in scientific database aims to optimize storing processes, retrieving and information maintenance through SMART (System for the Mechanical Analysis and Retrieval of Text) [51, 52]. In [5] it is proposed a basic structure for an Information Retrieval System (IRS) made by four elements: documents, user's search, how the elements are shown, and the evaluation function. The more extended IRS models on the internet are directories, search engines, and meta-searchers [53]. In literature, there are publications where the use of IRS is extended to general and particular contexts [1, 6, 54], giving as result search filters that can be found on a user's search module through the selection of relevant elements such as adequate attributes depending on which point of view is used:

- From the reader/researchers of scientific literature point of view where the proliferation of scientific journals and articles helps in the identification of the most relevant discoveries and research tendencies.
- From the authors and publishers point of view where papers and journals should generate indexes and abstract repertoire for a greater promotion of scientific production.

Search modules for editors, authors and researchers must base their results in metadata, which consequently is based in the recommendations made by most of meta-searchers "optimazers", taking into consideration the following elements [55]:

- NDF: Total number of found documents published by authors.
- HCD: Percentage of cited article.
- CD: Percentage document's cites without auto citations, related to the NDF cited.
- PDC: Percentage of cited documents.
- Hnf index: This index is calculated on a similar way to the h index but consider the different thematic cites categories and the number of authors per publication.
- NIR: Normalized indicators for researchers: this indicator gives the normalized average number of cites per document taking into account the different thematic cites categories where the researcher has published before.
- SNIPm: Median of the SNIP indicator for the set of journals where authors have published before.
- SJRm: median of the SJR indicator for the set of journals where authors have published before.
- Nm: Median of the number of authors per document. The Nm is a median of the normalized number of authors per document, taking into account the number of authors per thematic category.
- DIC: Percentage of documents with international collaboration.

4 Classic Search of Results

Classic searches are divided mainly by two factors: "on the site" that involves the element analysis that made the search such as Metadata Only and Full Text and Metadata which contains article title, abstract and indexation terms or both as default search. Within advance searching there are filters with the following labels: article title, authors, publication's title, abstract, indexation terms, authors affiliation, access number, article number, author's keywords, DOE terms, DOI, ISBN, ISSN, among others. The second factor, called "off the site" which involves quantity and quality study of the bibliographic citation in other articles, as well as the impact factor publishing journals and other data have that are external from scientific databases.

4.1 “on the Site” Results

The size of resources found through the search was under 30 KB of information. Moreover, regarding the attribute for a default research, concerning the prominence of keywords and the weight of the consulted word on the title we can observe that only 50 % of the article’s titles coincide with the search keyword. The latter makes reference to a tendency in the search on different databases, therefore the value may differ.

Another observed characteristic in the retrieved data is the lack of use of frames or the visual limitation of results due the size of the search. Most likely, this occurs because most of search engines have difficulties with process timing.

4.2 “off the Site” Results

This type of search does not belong to the site, but to the collected data by measurement external agents and publications’ quality. At this point the impact factor was taken, which is a numeric value assigned by external entities to scientific databases and it represents an important factor in the academic world.

5 Problem Description

As a partial conclusion we observed that there are keywords or attributes which determine the article positioning. Analyzing such process, it is inconclusive or difficult to make any value judgement regarding a search result and its relevancy. In fact, most of Searching Engine Optimization (SEO) tools indicate that by optimizing the off/on factors the real influence of such factors is unknown, even attributes as keywords can contribute to the introduction of “white noise” in the search made.

Knowing the relation between attributes, the weight each of them have and their temporal variability are complex factors that must be settled to achieve search criteria and article’s selection in obtaining positioning and thematic or temporal tendencies. The analysis of these factors may synthetize the manual construction of the state of the art.

Therefore, this article enhance a method of constructing a state of the art with semiautomatic learning techniques used in data mining which is capable to retrieve and synthetize information from attributes (keywords or key terms) to determine their behavior in any given temporal window.

6 J48 Classifier

J48 is an induction algorithm that generates a rule structure based on the building of a tree, which branches and leaves make part of data subsets extracted from the total set of data through a “training” phase. The algorithm generates a rule structure and evaluates its “kindness” while using criteria that measures the precision of said classification, which are:

- Value calculation of the information given by a candidate rule or branch of the tree through an "info" routine.
- A calculation of the global enhancement that said rule/branch by using a benefit routine.

With these two criteria, we calculated a cost-benefit function per each process cycle. This serves to decide if we can create a rule or if it is possible to merge rules into one.

It is worth mentioning that the output variable must be categorical and that this type of algorithms is based on information entropy. Therefore, the construction of a classification tree can be considered as an abstraction of order and disorder of data.

7 Construction Methodology State of the Art

The source for bibliographical data for scientific journals is more spread and accessible, resulting in a more complex analysis of content and new ideas association, that are the result of a relation between content and search, given that:

- Academic publications count with a database of 11.550 journal, from which there are 1,5 millions of articles per year [56].
- Publications regarding informatics behavior research based in text and social network information are slight irrelevant for a researcher and to cover filtered information and to select different sources of information turns to be a complex task. This mostly because the reading association analysis and subject selection is very demanding and time consuming.

Despite the keyword search reduces the amount of articles found, it does not mean the results obtained have any meaningful relation value regarding retrieve documents.

Data/text mining offers a solution to these problems, based in information retrieval techniques, natural language processing, information extraction and information regarding knowledge/discovery. A general model is illustrated in Fig. 1.

In phase 1, we retrieve enhanced information by consulting databases and filtered search through keywords, obtaining "relevant" articles.

In phase 2, we proceed with information preprocessing using keywords to lexicon analysis in such a way that the structured and unstructured data are extracted.

In phase 3, we applied a classification algorithms used commonly in data/text mining, which will allow the construction of a decision tree (J48) where each node represent keywords that serve to make partitions and connect to sub nodes between

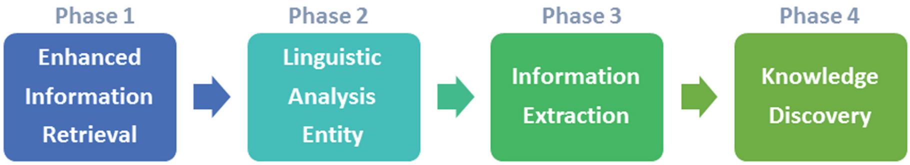


Fig. 1. Components of text mining

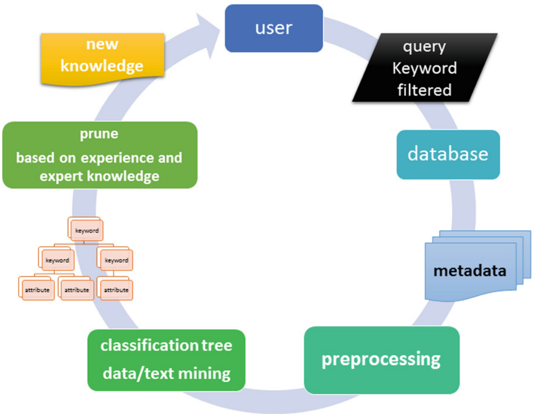


Fig. 2. Model data/text mining process

themselves, which represents sub values that may use academic databases attributes. The leaves represent subset cases that belong to each sub keyword. All paths from root to leaves create classification rules. As such, constructing the state of the art is made through supervised learning process with learning data, all the data found through filtered search of information in consulted databases.

In phase 4 we have a visualization and results analysis process. In this phase we extract useful information from different branches made in the classification tree. For this case study, we used the terms Internet of Things as roots, and leaves with terms such as sensors networks, IPv6, among others; extracted and visualized allowing to observe tendencies in a chronological order between several topics and sub keywords found. This information allows adding aggregated value at the moment of generating a state of the art, since this classification trees are based on entropy, in which their obtained results can show the order or disorder of the information.

The generated information to construct the tree can be explored to find pattern of new tendencies through reading of obtained results in phase 4 by an expert.

In Fig. 2, we show the process of a user to extract metadata from the academic databases.

8 Case Study Results: Internet of Things (IoT)

One of the tendencies that researchers have when it comes to write keywords is to put such keywords in order of relevancy. These keywords are associated with a weight with aim of having a classification parameter. At the moment of extracting keywords, some databases prioritize keywords in the title and abstract, getting data that does not correspond to indexation main keywords, which serve as a complementary data source for the search modules and in the application of advance filters in meta-searchers.

One of the consulting that compiles most of the synthetized content of an article is the cited bibliography. In most of the bibliographical data sources the BibText format is

Table 1. Format for processing

Keywords	Type of publication	Year	Weight
Internet of things	Journal	2014	1
Interoperability	Journal	2014	2
Self-organizing maps(SOM)	Journal	2014	3
Application agriculture	Proceedings	2013	1
⋮	⋮	⋮	⋮

used, which is used as input source in the preprocessing part, which through gawk filters or Excel data treatment that eliminates non relevant data to the consulted plain file. Data such as volume, month, and number of pages, among others are eliminated leaving the data visualized in Table 1.

There are thousands of keywords, so other intermediary processes take place before using the classification algorithm which consists in making these keywords discrete. In order to do that, the data in Table 1 are organized and they are associated to a topic assigning a label that identifies each group; the more itemized the associations are, the more quality the result will have. It is worth to mention that this classification must be done by a person that knows about the subject (expert).

With the discretized data and an adequate format for its processing, we run the J48 classification algorithm to filter the search. As the result of the search for Internet of Things (IoT), we get a classification tree shown in Fig. 3 and an extended version of the same in Fig. 4.

Figure 3 in relation with Fig. 4 we observed that groups are substituted by keywords they represented and repeated keywords have been removed leaving only the lesser depth level keyword. The resulting tree is made by a three year time frame and we observed that the research topics in Internet of Things have increased in recent years.

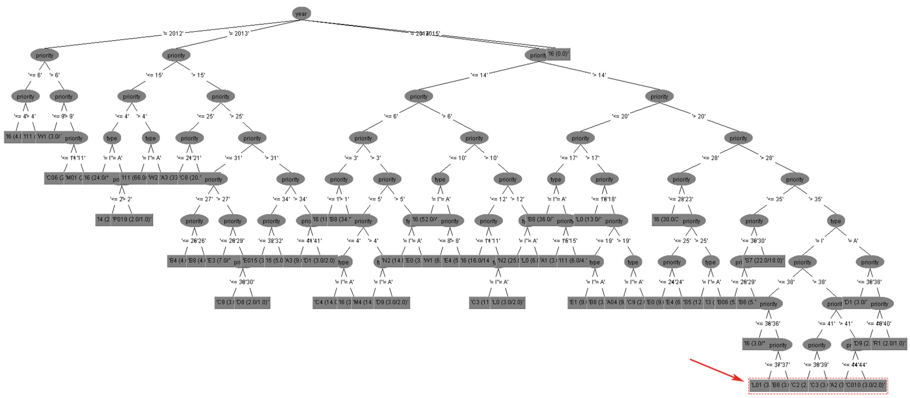


Fig. 3. Tree classification para internet of things

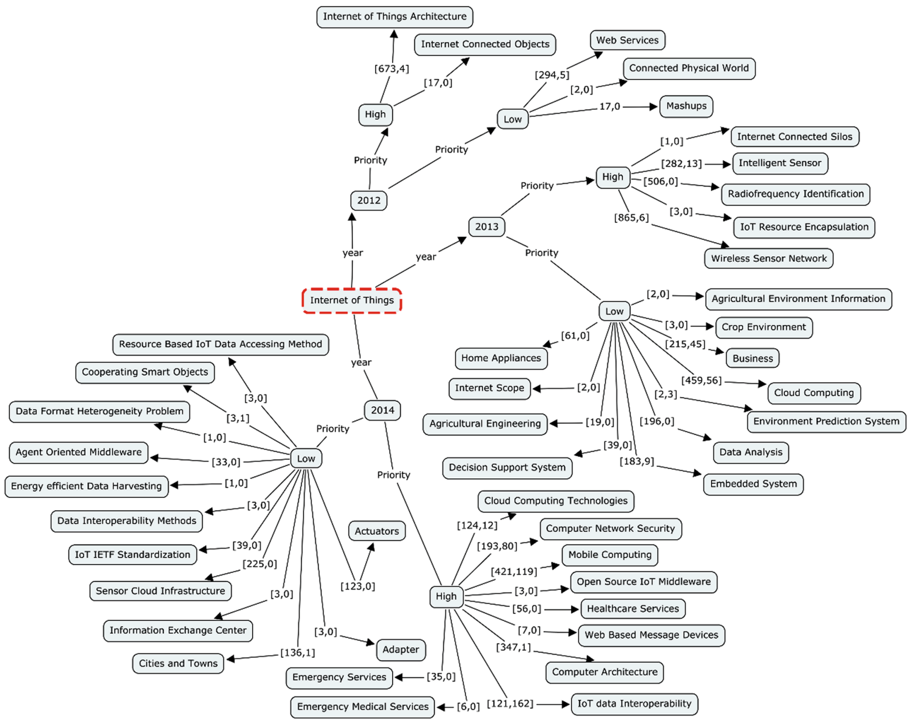


Fig. 4. State of the art, built with the J48 classifier

In Fig. 3, the deepest level in the tree is highlighted and a zoom in this section is shown in Fig. 5. This section represents more "disorganized" subjects which interpretation belongs to subjects least studied or published. For example, from a group (C2) the relevant keywords were "IoT Computational modeling", from group (C3) "IoT Computer architecture", from group (A2) "IoT middleware multiagent" and from (C010) "IoT Cooperating Smart Objects". According to tendency treated in [57] we added the word "Evolutionary" to the less treated subject, the obtained results were search in IEEE Xplore. ScienceDirect, and Google Scholar databases. These results were obtained on January 30th 2015 and they are shown in Table 2.

In Fig. 4 we observed that the tree that represents the state of the art for Internet of Things for the time frame 2012 to 2014. We added to this tree a descriptor that goes between brackets, i.e. for 2012 In the branch of lower priority, we find the keyword or subject "Web Services"; this subject is associated to data [294,5] in which the first value "294" represents the number of article that have associated that keyword somewhere in the article: document title, publication title, abstract, indexation terms, author's keywords or subtopic. Those previous subject are known as "indexation terms". The second value in that data "5" makes reference to the publications that has that keyword in the article title.

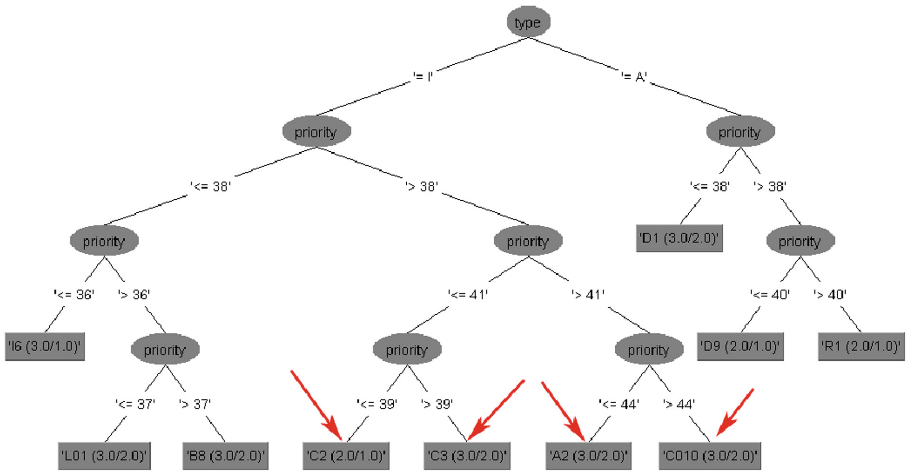


Fig. 5. Expansion of the least treated groups

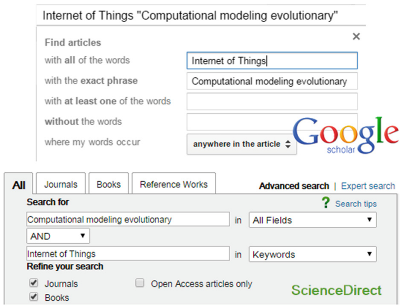


Fig. 6. Advanced search filters

Table 2. Thematic less explored in the internet of things

Keywords possible topics /areas of Innovation	Results databases		
	IEEE Xplore	ScienceDirect	Google scholar
IoT computational modeling evolutionary	1	11	4
IoT middleware multiagent	2	5	12
IoT cooperating smart objects	3	11	26
IoT computer architecture evolutionary	5	21	0

Once we made the search of the keyword in the tile, we are filtering the subject to a direct research, while if we search the keyword within the indexation terms, we get as result researches, directly or indirectly associated with the research subject.

The search filters used for ScienceDirect and Google Scholar databases is shown as it follows.

9 Conclusions

The methodology proposed in this paper gives general information regarding the development of a given subject within a time frame, locating the researcher in the identification of general subjects that have been studied and those subjects that have been less studied, because currently, academic and scientific databases do not possess visual tool that allow researchers to infer about the evolution and development of a general subject.

The decision/classification tree J48 allows us to see the thematic routes that have been studied within a time frame. The nodes closer to the root are the most used subject, also known as "indexation terms" that makes part of the article, either in the title, the abstract, author keywords, subtopics used direct or indirectly; because the process building the tree is based on entropy, sorting the information and the nodes or leaves located at the extremes or near the center of the tree belong to subjects that haven't been published extensively, becoming useful to contribute where there are gaps.

9.1 Future Research

For further research, we propose to compare this methodology based in J48 classifying algorithms with other classification algorithms. This with the aim of evaluating the relevancy of the Keywords grouping in the root node, being those subjects more studied.

The discretization can be automatized by using clustering algorithms aiming to automatize the process. We recommend integrating search engines from external companies such as Thomson Reuters, ISI, Scopus, Scimago, Google Scholar, Google Analytics, among other to the methodology. Also, we recommend that in each and every phase of this methodology requires extra work to enhance the data visualization and to find new ways to interpret the data.

References

1. Arduini, D., Zanfei, A.: An overview of scholarly research on public e-services? a meta-analysis of the literature. Telecomm. Policy **38**, 476–495 (2014). doi:10.1016/j.telpol.2013.10.007
2. Berendt, B., Krause, B., Kolbe-Nusser, S.: Intelligent scientific authoring tools: interactive data mining for constructive uses of citation networks. Inf. Process. Manag. **46**, 1–10 (2010). doi:10.1016/j.ipm.2009.08.002
3. Bornmann, L.: Do altmetrics point to the broader impact of research? an overview of benefits and disadvantages of altmetrics. J. Informetr. **8**, 895–903 (2014). doi:10.1016/j.joi.2014.09.005

4. Bornmann, L., Leydesdorff, L.: The validation of (advanced) bibliometric indicators through peer assessments: a comparative study using data from InCites and F1000. J. Informetr. **7**, 286–291 (2013). doi:10.1016/j.joi.2012.12.003
5. Bouyssou, D., Marchant, T.: An axiomatic approach to bibliometric rankings and indices. J. Informetr. **8**, 449–477 (2014). doi:10.1016/j.joi.2014.03.001
6. Bursa, M., Lhotska, L., Chudacek, V., et al.: Information retrieval from hospital information system: increasing effectivity using swarm intelligence. J. Appl. Log. (2014). doi:10.1016/j.jal.2014.11.006
7. Chakraborty, V., Chiu, V., Vasarhelyi, M.: Automatic classification of accounting literature. Int. J. Acc. Inf. Syst. **15**, 122–148 (2014). doi:10.1016/j.accinf.2014.01.001
8. Chang, S.-B., Lai, K.-K., Chang, S.-M.: Exploring technology diffusion and classification of business methods: using the patent citation network. Technol. Forecast. Soc. Change **76**, 107–117 (2009). doi:10.1016/j.techfore.2008.03.014
9. Cheng, B., Wang, M., Mørch, A.I., et al.: Research on e-learning in the workplace 2000–2012: a bibliometric analysis of the literature. Educ. Res. Rev. **11**, 56–72 (2014). doi:10.1016/j.edurev.2014.01.001
10. Cobo, M.J., Martínez, M.A., Gutiérrez-Salcedo, M., et al.: 25 years at knowledge-based systems: a bibliometric analysis. Knowl. Based Syst. (2015). doi:10.1016/j.knosys.2014.12.035
11. Coombes, P.H., Nicholson, J.D.: Business models and their relationship with marketing: a systematic literature review. Ind. Mark. Manag. **42**, 656–664 (2013). doi:10.1016/j.indmarman.2013.05.005
12. Dias, G.P.: Bibliometric analysis of portuguese research in e-government. Procedia Technol. **16**, 279–287 (2014). doi:10.1016/j.protcy.2014.10.093
13. Ding, Y., Fu, X.: The research of text mining based on self-organizing maps. Procedia Eng. **29**, 537–541 (2012). doi:10.1016/j.proeng.2011.12.757
14. Essendorfer, S., Diaz-Rainey, I., Falta, M.: Creative destruction in wall street's technological arms race: evidence from patent data. Technol. Forecast. Soc. Change (2015). doi:10.1016/j.techfore.2014.11.012
15. Fahimnia, B., Sarkis, J., Davarzani, H.: Green supply chain management: a review and bibliometric analysis. Int. J. Prod. Econ. (2015). doi:10.1016/j.ijpe.2015.01.003
16. Feng, F., Zhang, L., Du, Y., Wang, W.: Visualization and quantitative study in bibliographic databases: a case in the field of university–industry cooperation. J. Informetr. **9**, 118–134 (2015). doi:10.1016/j.joi.2014.11.009
17. Ferenhof, H.A., Vignochi, L., Selig, P.M., et al.: Environmental management systems in small and medium-sized enterprises: an analysis and systematic review. J. Clean. Prod. **74**, 44–53 (2014). doi:10.1016/j.jclepro.2014.03.027
18. Ferreira, M.P., Santos, J.C., de Almeida, M.I.R., Reis, N.R.: Mergers & acquisitions research: a bibliometric study of top strategy and international business journals, 1980–2010. J. Bus. Res. **67**, 2550–2558 (2014). doi:10.1016/j.jbusres.2014.03.015
19. Franceschet, M., Costantini, A.: The first Italian research assessment exercise: a bibliometric perspective. J. Informetr. **5**, 275–291 (2011). doi:10.1016/j.joi.2010.12.002
20. Franceschini, F., Maisano, D., Mastrogiacomo, L.: Scientific journal publishers and omitted citations in bibliometric databases: any relationship? J. Informetr. **8**, 751–765 (2014). doi:10.1016/j.joi.2014.07.003
21. Furukawa, T., Mori, K., Arino, K., et al.: Identifying the evolutionary process of emerging technologies: a chronological network analysis of world wide web conference sessions. Technol. Forecast. Soc. Change. **91**, 280–294 (2015). doi:10.1016/j.techfore.2014.03.013

22. Glenisson, P., Glänzel, W., Janssens, F., De, M.B.: Combining full text and bibliometric information in mapping scientific disciplines. Inf. Process. Manag. **41**, 1548–1572 (2005). doi:10.1016/j.ipm.2005.03.021
23. He, W.: Examining students online interaction in a live video streaming environment using data mining and text mining. Comput. Hum. Behav. **29**, 90–102 (2013). doi:10.1016/j.chb.2012.07.020
24. He, W.: Improving user experience with case-based reasoning systems using text mining and Web 2.0. Expert Syst. Appl. **40**, 500–507 (2013). doi:10.1016/j.eswa.2012.07.070
25. Huang, L., Zhang, Y., Guo, Y., et al.: Four dimensional science and technology planning: a new approach based on bibliometrics and technology roadmapping. Technol. Forecast. Soc. Change **81**, 39–48 (2014). doi:10.1016/j.techfore.2012.09.010
26. Huggett, S.: Journal bibliometrics indicators and citation ethics: a discussion of current issues. Atherosclerosis **230**, 275–277 (2013). doi:10.1016/j.atherosclerosis.2013.07.051
27. Janssens, F., Leta, J., Glänzel, W., De, M.B.: Towards mapping library and information science. Inf. Process. Manag. **42**, 1614–1642 (2006). doi:10.1016/j.ipm.2006.03.025
28. Jung, J.J.: Knowledge distribution via shared context between blog-based knowledge management systems: a case study of collaborative tagging. Expert Syst. Appl. **36**, 10627–10633 (2009). doi:10.1016/j.eswa.2009.02.052
29. Kostoff, R.N., Tshiteya, R., Pfeil, K.M., Humenik, J.A.: Electrochemical power text mining using bibliometrics and database tomography. J. Power Sources **110**, 163–176 (2002). doi:10.1016/S0378-7753(02)00233-1
30. Lee, C.-Y., Lee, Z.-J.: A novel algorithm applied to classify unbalanced data. Appl. Soft Comput. **12**, 2481–2485 (2012). doi:10.1016/j.asoc.2012.03.051
31. Leydesdorff, L., Bornmann, L., Marx, W., Milojević, S.: Referenced publication years spectroscopy applied to iMetrics: scientometrics, journal of informetrics, and a relevant subset of JASIST. J. Informetr. **8**, 162–174 (2014). doi:10.1016/j.joi.2013.11.006
32. Li, W., Zhao, Y.: Bibliometric analysis of global environmental assessment research in a 20-year period. Environ. Impact Assess. Rev. **50**, 158–166 (2015). doi:10.1016/j.eiar.2014.09.012
33. Lipitakis, A.D, Lipitakis, E.A.E.C.: On machine learning with imbalanced data and research quality evaluation methodologies. In: International Conference Computer Science and Computer Intelligence (CSCI) 2014, pp. 451–457 (2014)
34. López, V., Triguero, I., Carmona, C.J., et al.: Addressing imbalanced classification with instance generation techniques: IPADE-ID. Neurocomputing **126**, 15–28 (2014). doi:10.1016/j.neucom.2013.01.050
35. Mabe, M., Ware, M.: The STM report: an overview of scientific and scholarly journals publishing. STM (2009)
36. Martín-Valdivia, M.T., Díaz-Galiano, M.C., Montejo-Raez, A., Ureña-López, L.A.: Using information gain to improve multi-modal information retrieval systems. Inf. Process. Manag. **44**, 1146–1158 (2008). doi:10.1016/j.ipm.2007.09.014
37. McNay, I.: Research quality assessment. In: McGaw, P.P.B. (ed.) International Encyclopedia of Education, 3rd edn, pp. 307–315. Elsevier, Oxford (2010)
38. Mullins, M.E.: Has the time come for bibliometrics and the h-index in academic radiology? Acad. Radiol. **17**, 815–816 (2010). doi:10.1016/j.acra.2010.05.005
39. Nicholson, S.: The basis for bibliomining: Frameworks for bringing together usage-based data mining and bibliometrics through data warehousing in digital library services. Inf. Process. Manag. **42**, 785–804 (2006). doi:10.1016/j.ipm.2005.05.008
40. Ortega, J.L.: Relationship between altmetric and bibliometric indicators across academic social sites: the case of CSIC's members. J. Informetr. **9**, 39–49 (2015). doi:10.1016/j.joi.2014.11.004

41. Parvin, H., MirnabiBaboli, M., Alinejad-Rokny, H.: Proposing a classifier ensemble framework based on classifier selection and decision tree. Eng. Appl. Artif. Intell. **37**, 34–42 (2015). doi:10.1016/j.engappai.2014.08.005
42. Rafols, I., Leydesdorff, L., O'Hare, A., et al.: How journal rankings can suppress interdisciplinary research: a comparison between innovation studies and business & management. Res. Policy **41**, 1262–1282 (2012). doi:10.1016/j.respol.2012.03.015
43. Setsirichok, D., Piroonratana, T., Wongseree, W., et al.: Classification of complete blood count and haemoglobin typing data by a C4.5 decision tree, a naïve bayes classifier and a multilayer perceptron for thalassaemia screening. Biomed. Signal Process. Control **7**, 202–212 (2012). doi:10.1016/j.bspc.2011.03.007
44. Setti, G.: Bibliometric indicators: why do we need more than one? IEEE Access **1**, 232–246 (2013). doi:10.1109/ACCESS.2013.2261115
45. Shu, B., Kak, S.: A neural network-based intelligent metasearch engine. Inf. Sci. (Ny) **120**, 1–11 (1999). doi:10.1016/S0020-0255(99)00062-6
46. Taha, K.: Extracting various classes of data from biological text using the concept of existence dependency. IEEE J. Biomed. Health Info. 1, 1–15 (2015). doi: 10.1109/JBHI.2015.2392786
47. Triguero, I., Peralta, D., Bacardit, J., et al.: MRPR: a MapReduce solution for prototype reduction in big data classification. Neurocomputing **150**, 331–345 (2015). doi:10.1016/j.neucom.2014.04.078
48. Tsai, H.-H.: Global data mining: an empirical study of current trends, future forecasts and technology diffusions. Expert Syst. Appl. **39**, 8172–8181 (2012). doi:10.1016/j.eswa.2012.01.150
49. Tsai, M.-F., Wang, C.-J.: Post-modern portfolio theory for information retrieval. Procedia Comput. Sci. **13**, 80–85 (2012). doi:10.1016/j.procs.2012.09.116
50. Vassilakaki, E., Garoufallou, E.: Multilingual digital libraries: a review of issues in system-centered and user-centered studies, information retrieval and user behavior. Int. Inf. Libr. Rev. **45**, 3–19 (2013). doi:10.1016/j.iilr.2013.07.002
51. Vieira, E.S., Cabral, J.A.S., Gomes, J.A.N.F.: How good is a model based on bibliometric indicators in predicting the final decisions made by peers? J. Informetr. **8**, 390–405 (2014). doi:10.1016/j.joi.2014.01.012
52. Wagner, C.S., Roessner, J.D., Bobb, K., et al.: Approaches to understanding and measuring interdisciplinary scientific research (IDR): a review of the literature. J. Informetr. **5**, 14–26 (2011). doi:10.1016/j.joi.2010.06.004
53. Wang, B., Chai, K.-H., Subramanian, A.M.: Roots and development of intellectual property management research: a bibliometric review. World Pat. Inf. (2015). doi:10.1016/j.wpi.2014.12.002
54. Xie, X., Bai, Q., Hou, L., Wu, X.: Study and application of semantic-based image retrieval. J. China Univ. Posts Telecommun. **20**, 136–142 (2013). doi:10.1016/S1005-8885(13)60209-5
55. Xu, Z., Li, P., Wang, Y.: Text classifier based on an improved SVM decision tree. Phys. Procedia **33**, 1986–1991 (2012). doi:10.1016/j.phpro.2012.05.312
56. Yan, E.: Research dynamics: measuring the continuity and popularity of research topics. J. Informetr. **8**, 98–110 (2014). doi:10.1016/j.joi.2013.10.010
57. Dorronsoro, B., Ruiz, P., Danoy, G., et al.: Evolutionary Algorithms for Mobile Ad hoc Networks. Nature Inspired Computing Series. John Wiley & Sons, Hoboken (2014). doi:10.1002/9781118833209. To appear

Learning Ontology from Text: A *Storytelling Exploratory Case Study*

Valdemar T.F. Confort, Kate Revoredo, Fernanda Araujo Baião, and Flávia Maria Santoro(✉)

Programa de Pós-Graduação em Informática,
Departamento de Informática Aplicada,
Universidade Federal do Estado do Rio de Janeiro (UNIRIO),
Av. Pasteur, Rio de Janeiro, RJ, Brazil
{valdemar.confort, katerevoredo, fernanda.baiao, flavia.santoro}@uniriotec.br

Abstract. Business and IT systems are facing increasingly complex environments characterized by collaboration, change and variety of customers, suppliers and products. Appling group storytelling technique can contribute to the organization knowledge management. It brings benefits from capture to disseminating information, through communication and understanding of the concepts. American companies (3M and Apple), Japanese (Sony and Toshiba) and European (ClubMed and Océ) already use this approach in practice. On the other hand, the Ontology Engineering can contribute towards improving the quality of information and offer a solution to address knowledge management systematically. However, the specification and manually made of ontology management can be expensive, tedious, biased and pruned to error. Aiming to contribute with the management and quality of information, we explore the automatic learning of ontologies, which is an approach that extracts ontologies from data, both structured and unstructured (text). This work presents a proposal to extract an ontology from the tacit knowledge of those involved in the field. An exploratory study was able to get an ontology automatically from stories told by a group from a university department.

Keywords: Knowledge management · Group storytelling · Ontology learning

1 Introduction

Companies are increasingly dependent on their business networks, and to collaborate with other organizations, dealing with a greater variety of products, customers and organizational infrastructure [1, 2]. Knowledge Management can help organizations deal with this challenge from the perspective of information management and communication. Registering the information related to the execution of tasks is essential for knowledge management. This registration can be done through stories using the storytelling group technique [2].

However, it is also necessary to maintain the solid knowledge base and semantic ambiguities controlled as much as possible. In this context, the Ontology Engineering can contribute to improving the quality of conceptual models [3] that, in turn, might

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 477–491, 2015.
DOI: 10.1007/978-3-319-21009-4_37

contribute to the communication and exchange of factual information between the participants, whether people or systems. Still, the manual extraction of the concepts that make up an ontology can be expensive, tedious, predisposed to error, biased to its developer, inflexible and specific to the purpose that motivates its construction [4]. Automatic learning of ontologies from unstructured data is an alternative for handling the difficulties pointed out.

Given this scenario and aiming to contribute with the management and quality of information, our research question is: how to automatically specify an ontology from the tacit knowledge of those involved with a domain? By tacit knowledge we mean the kind of knowledge that is difficult to transfer to another person by means of writing it down or verbalizing it [5]. The proposed solution is to use group storytelling combined with text mining-based ontology learning. Thus, the stories are built collaboratively, and they are used to automatically specify an ontology.

This paper presents an exploratory study of the proposed solution. Besides this introduction, the paper is composed as follows: Sect. 2 presents the theoretical background including the concepts and techniques used in our proposal; Sect. 3 describes the experience; Sect. 4 presents the next steps and related work and Sect. 5 concludes the paper.

2 Theorectical Background

2.1 Group Storytelling

Storytelling, more specifically, Group Storytelling was the technique applied in our approach to collect knowledge from the domain people in order to have input to build the ontology. In his work, Tobin and Snyman [6] conducted a literature review on the use of stories in various organizations and points out that all organizations, large or small, are dependent of stories for its operation [7]. In the business context, the authors define the stories as sequence of decisions, actions or events (past, present or future; real or fictional) that involves characters in an organization where a business challenge or an opportunity should be addressed. They can be told in any form (written or oral) and use a variety of media. Storytelling is can be defined as the practices, tools and roles performed by those involved in communicating a story to an audience [6].

Tobin and Snyman [6] discuss the opinion of other authors who affirm a number of storytelling benefits: stories tend to be more well established that abstract ideas; enable understanding of a relevant process; is an effective way to capture knowledge; allows for quick, natural, clear, reliable, collaborative, persuasive, and accurate communication; allows the sharing of knowledge in a diverse population; is full of reason and emotion. They also refer to the practical use in American companies (3M and Apple), Japanese (Sony and Toshiba), European (Club Med and Océ), among others. Regarding group storytelling, there are still other benefits. Comments and communication are able to activate the memory and increase the ability of members to improve the narrative. In stories told by work teams, each participant can interact and present its own perception of the facts, contributing with other perspectives [8].

2.2 Conceptual Modeling and Ontologies

Conceptual modeling is "essentially concerned with the challenge of representing conceptual knowledge in a way that is appropriate and understandable to all stakeholders in the development and use of these representations, and independent of any eventual realization representation technique" [3]. Describing a system in terms of conceptual models consists of objects belonging to different classes, with different properties, and related with each other in various ways [1]. A generic approach to conceptually model a domain (universe of discourse) is the use of grammar and modeling methods to build a conceptual scheme [9]. Examples of grammars are the Entity Relationship (ER) model [10] and the Unified Modeling Language (UML) [11]; as an example of modeling methods, we can mention the Structured Analysis [12].

The Ontology Engineering amplifies the objectives beyond the creation and model specification. In this case, starting from a known universe of discourse, the process uses methods such as MethOntology, WebODE or On-To-Knowledge in conjunction with ontological languages such as OWL or RDF [9] apud [13–18]. Figure 1 shows the elements of these two approaches: modeling grammars combined with conceptual modeling methods producing conceptual schemes; and the Ontology Engineering. These approaches are not mutually exclusive; rather, they are complementary, the use of Ontology Engineering leads to better conceptual models and better conceptual models lead to information systems with higher quality.

In Philosophy, ontology is the basic description of things in the world [9]. Guarino [19] refines the philosophical meaning of ontology as a particular system of categories that reflect a particular view of the world. In the computer science and information systems field, ontology "refers to an engineering artifact, consisting of a specific vocabulary used to describe a certain reality, plus a set of explicit assumptions in accordance with the intended meaning of words vocabulary". The adoption of ontologies in conceptual modeling enables a higher level of reuse of knowledge, enabling reuse and sharing of domain knowledge between systems on heterogeneous platforms [19].

Ontology-driven Information Systems start with horizons personally performed (i.e., personal interpretations based on knowledge and experience). These horizons are

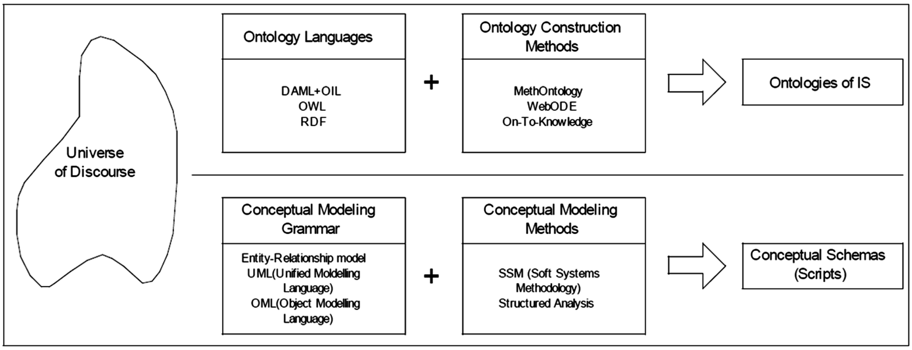


Fig. 1. Comparison of Ontology engineering and conceptual modeling [9].

Fig. 2. Path between personal horizon and the facts in the database [9].

described in some language, formal or informal. This ontology is, by definition, broader than the conceptual schemes and may include views of a larger community. They are theories that describe and explain the domain. Thus, the ontology provides a grammar for the conceptual model able to validate them in relation to reality [9]. Figure 2 summarizes the path between the personal horizon and the facts in the database. Therefore, we have chosen to build an ontology as a conceptual modelto represent the domain knowledge.

2.3 Ontology Learning

In this section, we present the concepts related to Ontology learning. They serve as a base to understand the method applied (discussed as follows). Ontology learning techniques aim to support the task of creating and maintaining an ontology [20]. It consists of a set of techniques to build from scratch, enhance or adapt an existing ontology automatically or semi-automatically using various resources [4] cited [21, 22]. Learning ontologies uses knowledge from various fields such as machine learning [23], natural language processing [24], information retrieval [25], artificial intelligence [26] and database management [4] cited in [22, 27]. It can be categorized according to the types of input data: structured, semi-structured and unstructured. Unstructured data can be text, such as books or magazines. Semi-structured data can be text in HTML or XML files. Structured data is data in databases. Our exploratory study focuses on unstructured data.

Ontology learning approaches based on unstructured data can be categorized as: statistical approach; natural language processing; mixed approach. The first approach is explained by Hazman El-Beltagy and Rafea [4], and generally, uses statistical measures to ensure the representation of a keyword in a number of texts and then further rounds are performed in the search of documents where the word was found to meet children of these concepts (representing words near them, for example) and thus determine an hierarchy.

The second approach uses a set of syntactic natural language patterns to discover the dependency relationship between words. In the working of Sabou et al. [27], for example, ontology extraction steps are: dependency analysis, syntactic patterns, ontology construction, pruning the ontology. As an example, Hazman et al. [4] cites the dependency relationship between the terms that can be recognized in the phrase “find antigenic sites on protein”. The “antigenic” adjective modifies the noun “site” and “site is the object of the verb ”find“. So, it is possible to annotate a text syntactically and, using a set of patterns, organize information to build an ontology. For example, Sabout et al. [27] use a syntactic pattern “NN” and “Nmod” to identify the concepts; verbs identify features; prepositions identify part-whole relationships between two terms (meronymy). An automated method using this approach is described by Cimiano, Hotho and Staab [28].

The third approach uses both semantic annotations and statistical analysis. This approach is applied in Text2Onto, a framework for ontology learning from textual resources. Figure 3 shows the architecture of Text2Onto which is based on the Probabilistic Ontology Model (POM). POM stores the result of different ontology learning algorithms. POM can be translated into ontology representation languages such as RDF, OWL and F-Logic. This ability is due to one Modeling Primitives Library (MPL, English, Primitive Modeling Library), which defines the primitives of languages declaratively and writers of ontologies, which are responsible for translating the primitives instantiated in one of knowledge representation languages. The algorithms are initialized by a controller (Controller Algorithm) which triggers a linguistic preprocessing of data (NLP, English, Natural Language Processing) in Corpus available. The controller executes the learning algorithm in the proper order in the reference repository (database accessed and controlled by references manager) and apply the necessary changes in the POM. The work of Cimiano and Volker [29], from which we extract Fig. 3 explains in greater detail the framework.

The scope of our contribution is an exploratory study using unstructured data obtained from the group storytelling technique. The objective was to evaluate the results obtained from the application of a method to automatically learn an ontology from the stories and thus contribute to the management and quality of information.

2.4 A Method for Automatic Extraction of Ontologies

There are several methods for building ontologies. We can cite the OTK [30], Methontology [18] and Diligent [31], for example. These three methods have the focus on engineers of ontologies and not machines and therefore they are not our focus, since we are interested in automatic approaches. The automatic approach we used in this paper is based on the Bedini & Nguyen [32], which is a process for automatic ontology

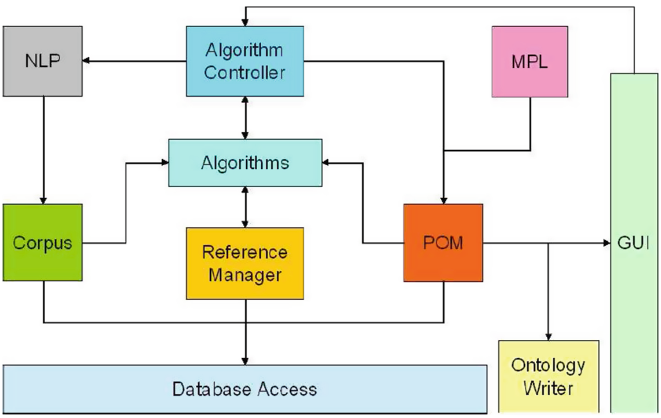


Fig. 3. Text2Onto architecture [29].

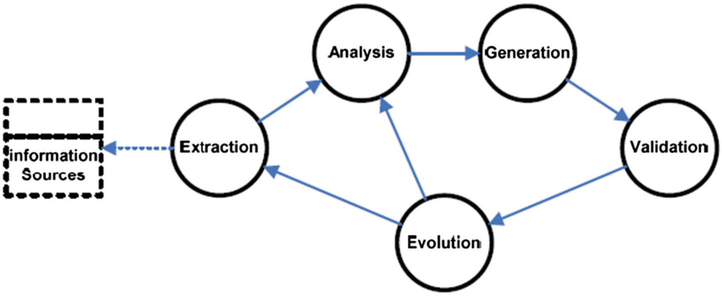


Fig. 4. The process of automatic generation of ontologies [32]

generation and evaluation criteria. Figure 4 depicts this process, described in detail as follows.

In the extraction step the information necessary for generating the ontology (concepts, attributes, relationships and axioms) is obtained from a corpus. As we have already mentioned, this information can be structured, semi-structured or unstructured. The extraction technique varies according to this source, such as natural language processing (NLP), clustering, machine learning, semantic, morphological, lexical, or a combination of them.

The analysis step focuses on the matching of the retrieved information and/or the combination of two or more of the existing ontologies, depending on the case. This step requires: beyond the first step techniques, semantic analyzers to detect homonyms synonyms and other relationships.

The generation is the step of merging ontologies and formalization of meta-model used by that tool can be interpreted by other applications. The formalization is the generation of OWL and RDF-S documents.

The validation step aims to remove misconceptions and relationships introduced in the previous steps. It is usually made manually, but in some cases can be automated.

The last step is evolution, since the domain is not static. There may be changes in the amount of concepts, relationships and other parameters can be added or modified.

3 Exploratory Study

The exploratory study of this research was carried out from the texts obtained from group storytelling[1] tool named CONTi. The tool allows stories to be told in asynchronous and distributed manner and recorded in texts. The dynamics is carried out in three stages. At first, the phase "Tell", the embers of the group participate telling the history, including, excluding and discussing the information provided by each. In the second phase, "Writing", history is reviewed by one member in order to be homogenously expressed, and, in the last phase, "Conclusion", the story is finished and can be commented.

[1] http://www.se7ti.com.br:8080/conti/login.

In this dynamic four roles are assigned. The participant contributes by telling the events that compose the history. The facilitator is responsible for observing and encouraging the group to ensure success. The editor is responsible for producing the final text, formatted and optimized. And finally, the critic who highlights elements exposed by participants' tacit knowledge.

The stories used in this research were created by groups of students in the context of the discipline Business Process Management at a public university where the study was conducted. Each student group was responsible for the telling one of the administrative processes, which they participate. From the texts of stories in the group storytelling tool, we explored the automatic learning of ontology concepts. We describe in detail the application of the learning ontology method in the next sub-sections.

3.1 Phase 0 - Planning

The main objective of the planning phase is to build a route that can guide the exploratory study. The corpus obtained through group storytelling was already available. We obtained it from a Business Process Management class by using a specific tool available to register stories using the group storytelling concepts. Since the stories were in Portuguese, the second step defined was the translation.

The third step planned was the concept extraction. In the planning phase, we conducted a comprehensive review of the literature to verify the availability solutions of ontologies automatically or semi-automatically generated.

Besides the method of Bedini and Nguyen [32] and the Text2Onto framework of Cimiano and Volker [29], we have investigated other proposals in order to define our approach. The approach proposed by OntoLT [33] has been tried; in this case, the linguistic annotation function is external, and then it should be performed before using the tool for the extraction of ontology. We have to annotate a text semantically using the GATE[2] and tooling available at Stanford NLP Group,[3] but they were incompatible with OntoLT, and this would represent a big effort to make them compatible with the tool. Besides, OntoLT is a plugin for Protégé that only works with its older versions.

We also evaluated the use of OntoGen. This is a semiautomatic editor that combines mining techniques with machine learning. It is able to perform supervised and unsupervised learning to suggest concepts and names of concepts, as well as the visualization of the ontology [34]. However, it only supports the ontology management; it is not able to build it automatically.

Finally, we worked with the Text2Onto tool. It also presented difficulties, even in the installation phase. There are a number of incompatibilities with versions of other tools, but it proved able to extract the constructs fully automatically. Thus, we will present this evaluation.

---

[2] GATE – General Archictecture Text Engineering (https://gate.ac.uk/).

[3] http://nlp.stanford.edu/.

Fig. 5. Research approach for the exploratory study

A step to evaluate the extraction was also a result of the planning phase. In this research, we conducted a manual evaluation. As a separated step, it can be evolved in a more specific research.

The visualization of the result is the last step. Through the visualization, the ontology can be not only analyzed but also edited. This phase can be evolved in a more specific research as well.

Thus, our approach follows, beyond this planning step, four steps. In the second step, we conducted the pre-processing of our data source. In the third step we perform the automatic extraction of concepts. In the fourth step, we evaluate the results compared to manual extraction. In the last step, we display the concepts extracted within an ontology editing tool. Figure 5 depicts this research approach

4 Phase 1 – Pre-processing (Translation)

The extractions of the histories both in Portuguese and English are available online.[4] The processes were grouped into different histories (see Fig. 6, which displays the group storytelling tool interface). Each history was then been translated into English and included in a separate document. All the documents, in English, composed the corpus of this work. The translation was performed using Google Translate.[5]

For instance, there is a group story about the breach of requirements of a discipline in a university. It was originally in Portuguese. An example of the text in Portuguese is:

> *"Esse processo acontece quando um aluno quer fazer uma matéria que depende que outra seja concluída antes, mas ele ainda não a fez."*

Then we copy and paste the text, already finished by the group, to a tool that could perform the translation. Using the Google Translate the result obtained for the example above was:

> *"This process happens when a student wants to do a story that depends on another to complete before, but he still did it."*

Although we know that translations tools are not a hundred per cent reliable, for next phase, we processed the raw text obtained by the translation process. No adaptation or correction was made in this exploratory research to avoid introducing a bias.

4 https://www.dropbox.com/sh/0gpkqbo5tpggh38/AADGXaRDcfcd6jMxAv8M5CdKa?dl=0.
5 https://translate.google.com/.

Fig. 6. Histories available within Conti.

4.1 Phase 2 – Concept Extraction

We decided to work only on the concept level and defined in advance the algorithms to be used. The Text2Onto uses the notes taken in GATE[6] for which nouns are "concepts" and proper names (proper nouns) are "instances".

Since this research is exploratory we only evaluated the options available by default in the tool. For the extraction of concepts and instances, we used the TFIDFConceptExtraction algorithm. The other options available were the EntropyConceptExtraction; ExampleConceptExtraction; RTFConceptExtraction. This choice is justified by the fact that the ExampleConceptExtraction is an example algorithm; The RTF assesses the frequency of the term in the document; the TFIDF uses RTF, but makes its assessment also with respect to a list of documents and not for a single document [29].

We performed the extraction of concepts within a single document. Prior to the extraction, however, the concepts were extracted manually for comparison with the final result. We randomly choose the "DeclarationOfCompletionOfCourse" document. In the next section we present and evaluate the results.

4.2 Phase 3 – Evaluation

Table 1 presents the concepts extracted both manually and automatically. The third column is our analysis whether the correct extraction was performed (yes) or not (not) comparing the automatic and manual extractions.

Despite any statistical analysis, we found that the tool hit in 16 of 23 results. Looking at the 7 cases in error, we observe that all errors are the result of the frequent

[6] GATE – General Archictecture Text Engineering (https://gate.ac.uk/).

Table 1. Comparison of automatic and manual extraction of concepts.

Concepts extracted manually	Concepts extracted automatically	Analysis of the results
Application	Application	Yes
application form	application form	Yes
	Completion	No
Course	Course	Yes
Data	Datum	Yes
Date	Date	Yes
Declaration	Declaration	Yes
Degree	Degree	Yes
	department verify	No
Department		No
Direction	Direction	Yes
Document	Document	Yes
	document delivery	Yes
End	End	Yes
	Order	Yes
Request	Request	Yes
Secretariat	secretariat	Yes
Secretary	Secretary	Yes
	Sign	No
Stamp	Stamp	Yes
Statement	Statement	Yes
	statement verify	No
Student	Student	Yes

disambiguation problem been reported in several studies of semantic analysis and which is currently a challenge for the area [35]. With this result, we believe that the extraction of concepts was satisfactory. However, we conclude that the method needs to be revised to include a new phase, in order to reduce the ambiguities. In this sense, the work of Leão, Revoredo and Baiao [35] could provide a basis to support this new phase.

4.3 Phase 4 – Visualization

At this stage the goal is to display the concepts extracted using ontology management tools. Although the export functionality for OWL does not work, it performed well the exportation to RDF. Then we visualized the ontology in the most frequently used tools: Protects and NeonToolkit.[7] The results can be seen in Figs. 7 and 8.

This phase is important to make sure if there is adhesion between the results obtained from the learning tools and construction/maintenance of ontologies tools. We concluded that the ontology generated by this approach can be used within the tools

[7] http://neon-toolkit.org/wiki/Main_Page.

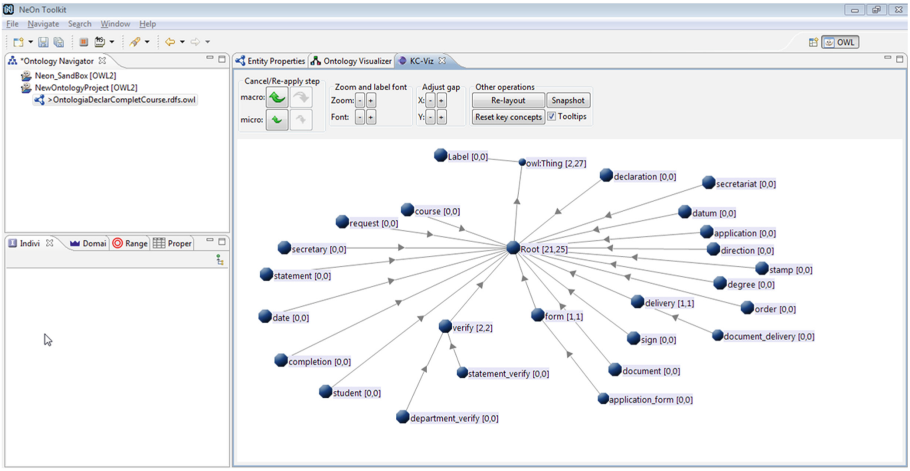


Fig. 7. Concepts displayed in NeonToolkit

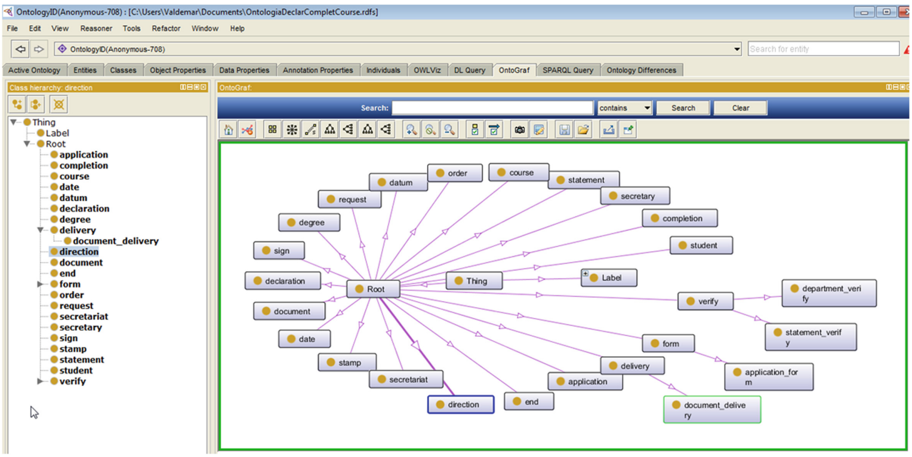


Fig. 8. Concepts displayed in the Protege.

available to criticize it or evolve it. This is probably due to the maturity of this tooling, including OWL and RDF standards. These steps (ontology construction and sedimentation of patterns) above, both in terms of research and in practice, precede the evolution of ontology learning step.

5 Related Work and Next Steps

The first step after the completion of this exploratory study is to extend it. In this work we presented a step by step which allowed the extraction of concepts from one of the stories available on the group storytelling tool. It can be extended it in five aspects.

At first, in relation to the stories, we can repeat the same approach for other stories available, composing an ontology from all the stories. Mining processes from group storytelling have been already conducted [36].

The second aspect is related to the ontological elements. We only assessed concepts (classes). The approach can be extended to obtain instances, relationships and categorizations dealing with other kinds of challenges. The literature is vast in this direction and we have cited here some works, such as: Cimiano, Hotho and Staab [28], Staab and Studer [20], Shamsfard and Barforoush [21].

The third aspect is the evaluation. Human review is not the only way to check the results. Ontology matching verification techniques can be one way. Several studies aim to evaluate the similarity in order to align ontologies. One example is presented by Jain et al. [37], where it is proposed a system called BLOOMS that is capable of evaluating ontology alignment.

The fourth aspect is the comparison with other languages. Works extending the ontology learning for other languages can be cited in Chinese [38] or in Portuguese [39]. Cross-linguistic work can also be cited [40, 41].

The fifth aspect is related to foundational ontologies [42]. Leão, Revoredo and Baiao [35] intends to obtain a well-founded ontology using natural language processing coupled with the Wordnet elements. In this case we might generate an ontology less ambiguous or subjected to diverse interpretations.

6 Conclusion

We conclude this work emphasizing the importance and relevance of the topic. In this work we were able to contribute to answer our research question: "how to automatically specify an ontology from the tacit knowledge of those involved with a domain?"

We proposed an approach in which in the first phase the tacit knowledge of those involved in the field is captured collaboratively using group storytelling technique. In the second phase, we are able to learn the concepts needed to build an ontology automatically.

Although in 2007, the work of Bedini & Nguyen [32] presented the state of the art automatic generation of ontologies and the authors concluded that there was no system capable of automatically generating ontologies; our exploratory study on the contrary provided some evidence of this possibility. However, we also found that the tools for automatic ontology learning from texts, perhaps by the strong dependence of semantic annotation, have not reached yet a state of the art enabling the ease of use.

As a result, future work will be advance the extraction of ontological elements. This research indicates that, in this sense, we should also improve the code of the tools available. This perception is mainly due to the verification of discontinuity of platforms which we exploited. For example, the Text2Onto uses only the, version 4 of GATE. The GATE, however, is at version 8.

Finally, another improvement is the comparison of the ontologies generated from descriptions obtained from the individual and collaboratively form. Although the advantages of obtaining knowledge from collaborative approaches are known, we found no references that make that comparison. Thus, we conclude that, in light of the results, the trend in this research is a further experimental approach and expand the approach to other ontological elements.

Acknowledgment. Fernanda Araujo Baião and Flávia Maria Santoro are partially funded by the CNPq brazilian research council, respectively under the projects 309069/2013-0 and 473457/2013-0. This work was partially funded by FAPERJ (project number E-26/110.477/2014).

References

1. Johannesson, P.: The role of business models in enterprise modelling. In: Krogstie, J., Opdahl, A.L., Brinkkemper, S. (eds.) Conceptual Modelling in Information Systems Engineering, pp. 123–140. Springer, Heidelberg (2007)
2. Perret, R., Borges, M.R., Santoro, F.M.: Applying group storytelling in knowledge management. In: de Vreede, G.-J., Guerrero, L.A., Marin Raventós, G. (eds.) CRIWG 2004. LNCS, vol. 3198, pp. 34–41. Springer, Heidelberg (2004)
3. Recker, J.: Evaluations of Process Modeling Grammars, vol. 71. Springer, Heidelberg (2011)
4. Hazman, M., El-Beltagy, S.R., Rafea, A.: A survey of ontology learning approaches. Int. J. Comput. Appl. **22**, 0975–8887 (2011)
5. Nonaka, I., Takeuchi, H.: The Knowledge Creating Company: How Japanese Companies Create the Dynamics of Innovation, p. 284. Oxford University Press, New York (1995)
6. Tobin, P., Snyman, R.: Once upon a time in Africa: a case study of storytelling for knowledge sharing. Aslib Proc. **60**(2), 130–142 (2008)
7. Mitroff, I., Kilmann, R.: Store managers tell: a new tool for organizational problem solving. Manag. Rev. **64**(7), 18–28 (1975)
8. Santoro, F.M., Brezillon, P.: The role of shared context in group. Comput. Inform. **25**(6), 1001–1026 (2006)
9. Fonseca, F.: The double role of ontologies in information science research. J. Am. Soc. Inform. Sci. Technol. **58**(6), 786–793 (2007)
10. Chen, P.P.-S.: The entity-relationship model - toward a unified view of data. ACM Trans. Database Syst. (TODS) **1**, 9–36 (1976)
11. Rumbaugh, J., Jacobson, I., Booch, G.: The Unified Modeling Language Reference Manual. Pearson Higher Education, Delhi (2004)
12. Yourdon, E.: Modern Structured Analysis, pp. 78–94. Prentice Hall, Saddle River, NJ (1989)
13. Fensel, D.: Ontologies: a silver bullet for knowledge management and electronic commerce. Springer, New York (2001)
14. Fensel, D.: Language standardization for the semantic web: the long way from OIL to OWL. In: Plaice, J., Kropf, P.G., Schulthess, P., Slonim, J. (eds.) DCW 2002. LNCS, vol. 2468, pp. 215–227. Springer, Heidelberg (2002)
15. Fensel, D., van Harmelen, F., Horrocks, I., McGuinness, D., Patel-Schneider, P.: OIL: an ontology infrastructure for the semantic web. IEEE Intell. Syst. **16**, 38–45 (2001)
16. Gomez-Perez, A., Benjamins, R.: Knowledge engineering and knowledge management: ontoligies and the semantic web. In: Benjamins, V.R. (ed.) 13th International Conference EKAW 2002. Springer, Heidelberg (2002)
17. Noy, N., Hafner, C.: The state of art in ontology Design. AI Mag. **18**, 53–74 (1997)

18. Gómez-Pérez, A., Fernández-López, M., Corcho, O.: Ontology Engineering. Springer, New York (2003)
19. Guarino, N.: Formal ontology in information systems. In: Proceedings of the First Internation Conference (FOIS 1998), vol. 46, 6–8 June 1998
20. Staab, S., Studer, R.: Handbook on Ontologies. Springer, New York (2010)
21. Shamsfard, M., Barforoush, A.A.: The state of the art in ontology learning: a framework for comparison. Knowl. Eng. Rev. **18**(4), 293–316 (2003)
22. Gómez-Pérez, A., Manzano-Macho, D.: A survey of ontology learning methods and techniques. Ontoweb Deliverable D1 (2003)
23. Mitchell, T.M., Carbonell, J.G., Michalski, R.S.: Machine Learning: A Guide to Current Research. Kluwer Academic Publishers, Norwell (2011)
24. Manning, C.D., Schutze, H.: Foundations of Statistical Natural Language Processing. The MIT Press, Cambridge (1999)
25. Manning, C.D., Raghavan, P., Schutze, H.: Introduction to Information Retrieval. Cambridge University Press, Cambridge (2008)
26. Russel, S., Norvig, P.: Artificial Intelligence: A Modern Approach, 3rd edn. Prentice Hall, Saddle River (2009)
27. Sabou, M., Wroe, C., Goble, C., Mishne, G.: learning domain ontologies for web services descriptions: an experiment in bioinformatics. In: Proceedings of the 14th International Conference on World Wide Web, Chiba (2005)
28. Cimiano, P., Hotho, A., Staab, S.: Learning concept hierarchies from text corpora using formal concept analysis. JAIR J. AI Res. **24**, 305–339 (2005)
29. Cimiano, P., Völker, J.: Text2Onto. In: Montoyo, A., Muńoz, R., Métais, E. (eds.) NLDB 2005. LNCS, vol. 3513, pp. 227–238. Springer, Heidelberg (2005)
30. Staab, S., Schnurr, H., Studer, R., Sure, Y.: Knowledge processes and ontologies. IEEE Intell. Syst. **1**, 26–34 (2001)
31. Tempich, C., Pinto, H., Sure, Y., Staab, S.: An argumentation ontology for distributed, loosely-controlled and evolving engineering processes of ontologies (DILIGENT). In: Gómez-Pérez, A., Euzenat, J. (eds.) ESWC 2005. LNCS, vol. 3532, pp. 241–256. Springer, Heidelberg (2005)
32. Bedini, I., Nguyen, B.: Automatic ontology generation: State of the Art (2007)
33. Buitelaar, P., Olejnik, D., Sintek, M.: A protégé plug-in for ontology extraction from text based on linguistic analysis. In: Bussler, C.J., Davies, J., Fensel, D., Studer, R. (eds.) ESWS 2004. LNCS, vol. 3053, pp. 31–44. Springer, Heidelberg (2004)
34. Fortuna, B., Grobelnik, M., Mladenic, D.: OntoGen: semi-automatic ontology editor. In: Smith, M.J., Salvendy, G. (eds.) HCII 2007. LNCS, vol. 4558, pp. 309–318. Springer, Heidelberg (2007)
35. Leão, F., Revoredo, K., Baião, F.: Learning well-founded ontologies through word sense disambiguation. In: Brazilian Conference on Intelligent Systems, Fortaleza - CE (2013)
36. Gonçalvez, J., Santoro, F., Baião, F.A.: Business process mining from group stories. In: Proceedings 13th International Conference on Computer Supported Cooperative Work in Design, Santiago, Chile (2009)
37. Jain, P., Hitzler, P., Sheth, A.P., Verma, K., Yeh, P.Z.: Ontology alignment for linked open data. In: Patel-Schneider, P.F., Pan, Y., Hitzler, P., Mika, P., Zhang, L., Pan, J.Z., Horrocks, I., Glimm, B. (eds.) ISWC 2010, Part I. LNCS, vol. 6496, pp. 402–417. Springer, Heidelberg (2010)
38. Wen, C., Wang, X.-B., Shi, Z.-x.: Automatic domain-specific term extraction in Chinese domain ontology learning. Appl. Res. Comput. **7**, 73 (2009)

39. Vieira, R., Ribeiro, L.: Ontolp: Engenharia de ontologias em língua portuguesa. In: *Anais do xxviii congresso da SBC-SEMISH-Seminário integrado de software e hardware*, Belem do Pará (2008)
40. Vossen, P.: EuroWordNet: A Multilingual Database with Lexical Semantic Networks. Kluwer Academic Publishers, Dordrecht (1998)
41. Leopold, H., Eid-Sabbagh, R.-H., Mendling, J., Azevedo, L.G., Baião, F.A.: Detection of naming convention violations in process models for differente languages. Decis. Support Syst. **56**, 310–325 (2013)
42. Guizzardi, G., Falbo, R., Guizzardi, R.: A importância de ontologias de fundamentação para a engenharia de ontologias de domínio: o caso do domínio de processos de software. IEEE Am. Lat. **6**, 244–251 (2008)

Knowledge Management and the Internet of Things

The Effects of the Internet of Things and Big Data to Organizations and Their Knowledge Management Practices

Jari Kaivo-oja[1(✉)], Petri Virtanen[2], Harri Jalonen[3], and Jari Stenvall[2]

[1] Finland Futures Research Centre, Turku School of Economics, University of Turku, Turku, Finland
jari.kaivo-oja@utu.fi
[2] School of Management, University of Tampere, Tampere, Finland
[3] Turku University of Applied Sciences, Turku, Finland

Abstract. New technologies are promising us many upsides like enhanced health, convenience, productivity, safety, and more useful data, information and knowledge for people and organizations. The potential downsides are challenges to personal privacy, over-hyped expectations, increasing technological complexity that boggles us. Our point is this change requires scientific discussion from the point of management, leadership and organizations – that is, it is time to discuss the meaning of these challenges seriously also in terms of existing traditions of management science. This review type article discusses the nature and role of the Internet of Things (IoT), Big Data and other key technological waves of ubiquitous revolution vis-á-vis the existing knowledge on management, organizations and knowledge management practices in organizations. Recent changes in the fields of robotics, artificial intelligence and automation technology indicate that all kinds of intelligence and smartness are increasing and organizational cultures are going to change indicating fast changes in the field of modern management and management sciences. Organizational processes form the base for the knowledge-based decision-making. Developing and utilizing smart solutions – like the utilization of Big Data – emphasize the importance of open system thinking. Digitalized services can for instance create new interfaces between service providers and users. Service users create social value while they are participating in co-producing activities. Hence, the IoT and Big Data undoubtedly strengthen the role of participation in service production, service economy, innovativeness in-between organizations (as a joint processes) and leadership models incorporated in service-dominant –logic. Moreover, IoT, Big Data, and especially digitalization bring about the renaissance of knowledge in decision-making.

At organizational level, smart organizations do no rely on knowledge production, but focus on knowledge integration instead. Knowledge integration becomes a key part of management systems. This also means that seminal theories with regard to decision-making and knowledge management do not suffice anymore. At organizational level there is a growing need to develop abilities to act in changing, not easy to forecasted and non-linear situations due to the complexity related to utilization and developing digitalization. Authentic and clinical leadership involves components such as awareness, unbiased processing, action, and relations. IoT and Big Data certainly effect organizations. The connection between IoT, Big Data, management systems as well as

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 495–513, 2015.
DOI: 10.1007/978-3-319-21009-4_38

knowledge management practices at organizational level has not been analysed thoroughly in the KMO theories or empirically so far. In this scientific article this task and challenge will be performed.

Keywords: The internet of things · Internet of intelligent things · Big data · Management · Leadership · Knowledge management

1 Introduction

New technologies are promising us many upsides like enhanced health, convenience, productivity, safety, and more useful data, information and knowledge for people and organizations. The potential downsides are challenges to personal privacy, over-hyped expectations, increasing technological complexity that boggles us. Technological complexity equals also with technology risks – no wonder then that there has been a growing discussion among the social scientists since over 20 years about the risks of the modern society [1]. Seemingly, new technologies always involve a fundamental paradox – i.e. it is simultaneously both a solution and a problem (cf. [2]). The paradox arises, for example, from the fact that while new technologies expands the information pool from which to draw decisions, they also simultaneously generate contradictory information that may make it difficult to achieve consensus.

We believe that the prevailing technological r/evolution changes organizations and institutions. Consequently, the relative importance of networks and crowds will increase in relation to hierarchies and markets. If technologies are used wisely, most people will be better off, but if technologies are used without smartness the results will be messy and even disastrous. Our point is in this conceptual paper that this change requires scientific discussion from the point of management, leadership and organizations – that is, it is time to discuss the meaning of these challenges seriously also in terms of existing traditions of management science. This article discusses the nature and role of the Internet of Things (later on referred as IoT), Big Data and other key technological waves of ubiquitous revolution vis-á-vis the existing knowledge on management, organizations and knowledge management practices in organizations. Recent changes in the fields of robotics, artificial intelligence and automation technology indicate that all kinds of intelligence and smartness are increasing and organizational cultures are going to change indicating fast changes in the field of modern management and management sciences. This paper is based on the literature survey in our previous work [3] and it develops further our main ideas with regard to public service systems, systemic change factors and the need to re-think current theories of change management, well-being at work, theories of motivation.

Conceptually speaking, the IoT refers to uniquely identifiable objects (things) and their virtual representations in an Internet structure.[1] Moreover, the IoT refers to

[1] The basic concept of IoT was initially applied in the Radio-Frequency Identification RFID-tags to mark the Electric Product Code (Auto ID-Lab). The benefit of IoT concept it that it enables physical objects to be seamlessly integrated into the information network, where the physical objects can become active participants in business and management processes (see e.g. [7]).

intelligent devices that have adequate computing capacity. IoT can also include relatively unintelligent sensors which may well feed into smarter devices. With regard to Big Data then, Boyd and Crawford [4] have argued that the era of Big Data has begun. Computer scientists, physicists, economists, mathematicians, political scientists, bio-informaticists, sociologists, and other scholars are simply clamoring for access to the massive quantities of information produced by and about people, things, and their interactions. To be more specific, the concept of Big Data is definitely very vague. For instance, Zalavsky et al. [5] point out that there is no clear definition for 'Big Data'-whereas it does not mean the only size of the data *per se*, it can be described with three characteristics with regard to data, as known as the three V's – volume, variety, and velocity of the data. More practically, Mayer-Schönberger and Cukier [6, p. 7] refer to Big Data as follows: big data refers "…to things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new fors of value, in ways which change markets, organizations, the relationship between citizens and governments and more".

This paper is organized as follows. First, we discuss the future of Internet, it's lates and the next development phase with special emphasis on the IoT. Secondly, we re-think the key issues of ubiquitous r/evolution and argue that these phenomena will have an impact on a multitude scale to e-business as well as to human interaction. Thirdly, we elaborate the impacts of all mentioned changes to knowledge management and knowledge-based decision-making. Fourthly, we analyze the connection between IoT, Big Data r/evolution, smart organizations with the idea to pinpoint what kind of societal impacts ubiquitous r/evolution and associated other changes pursue.

2 From Internet Rush to Internet Saturation

In 1991, Weiser [8] (1991) described the vision of the future world under the name of *Ubiquitous Computing*. Today, we have smart phones, our cars are computer systems on wheels, and our homes are turning into smart living environments. Also our production and delivery systems have developed to smartness and technology solution are really ubiquitous. SmartFactory and SmartService concepts are already ready for customers and citizens. In the 1980s electrical engineering was key engineering paradigm. In the 1990s, the key paradigm was software and mechanical engineering, which emphasized modeling object, systems and mechatronic objects and systems. After 2000 the key paradigm turned to *useware engineering*, which underlines modeling interactions [9].

Today key issues in e-commerce and e-trade are new applications and contexts, where applications are used. Today´s engineering systems are simple and not too complex, they are not based on centralized hierarchies, they allow for a really concurrent engineering by decoupling process, mechanical, electrical, and control design on the basis of semantic models, they create and apply standards to all levels of automation pyramid in order to reduce planning effort and allow re-use of components, and introduce technologies and applications for the human being and organizations [9, p. 137]. As regards to the Internet, it really has introduced the new institutional revolution in the globe. As a consequence, the importance of networks and crowds has increased in relation to market institutions. In the future their importance probably increases because of increasing coverage of Internet broadband and networks and the size of population using things and "gadgets" related to the Internet increases.[2] Summing up, the key variables in the Internet/r/evolution are population size (N), the resources feeding supply sub-system (F) encompassing both the natural resources and food system, and accumulated technological and scientific knowledge sub-system (K). This N-F-K triangle base is key platform for Internet r/evolution, human development and economic growth [10, pp. 744–745, pp. 750–751].

IoT is a part of latest information infrastructure with cloud computing environments, ubiquitous networks, Linked Data Web and autonomous decentralized systems. These new technologies provide both computing and communication quantifiable resources that offer flexible levels of business performance and quality of demand [11, pp. 159–160]. Obviously IoT will have huge economic and social impacts on business and public sector management and administration. We´ll discuss these effects later on in this paper. It appears that the pre-conditions of data and knowledge management will change fundamentally before 2040. A key aspect of this development process will be IoT. Also social and economic preconditions for crowdsourcing, Big Data and networking are much stronger than today by year 2040, when Internet saturation phase will be reached. This means that the so called Internet rush era is turning to Internet saturation era. Internet and its architectural principles were designed in the 1970s, in the beginning 5th Kondratieff cycle. Now we are starting 6th Kondratieff cycle, where new technologies of the Internet will be adopted. Recent analyses reveal that the importance of the Internet for human society is constantly increasing (e.g. [12, 13]). In last four decades the Internet has moved from being a restricted network of computer science researchers into being the global infrastructure of service economy and information society.[3]

[2] Internet penetration index will probably increase in the world in next decades. In 1990 the number of internet hosts was were small, but today the number of internet hosts is considerable and increasing. Current forecasts indicate that internet penetration will slow down towards saturation at about 89 % by the year 2040. It will take 45 years for the Internet penetration process to be globally accomplished. Experts estimate that average time lag of development phases of Internet is a four years and there months. It is predicted that in 2015 the penetration index should reach 50 % of the world population. From then on it should asymptotically approach saturation at 79-89 % by 2035-2040 (see e.g. [10]).

[3] To take an example: today over one billion people use it to communicate, search and share information, conduct business and enjoy entertainment. In 2010, online retail sale volume was estimated to be $697.8 billion in 2012 in the world. Thus, online retail has become an emergent trend throughout the world and the e-commerce industry continues to grow significantly across the world. Such issues like importance of convenience, personal innovativeness, impulsiveness, price consciousness, risk aversion, brand consciousness, variety seeking, attitudes towards online shopping and online advertising are key issues in the e-commerce market [12, 13]. In 2014 retail sales worldwide—including both in-store and internet purchases—will reach $22.492 trillion, according to new figures and statistics from eMarketer. The global retail market will see steady growth over the next few years, and in 2018, worldwide retail sales will increase 5.5 % to reach $28.300 trillion - See more at [14].

3 The Context: IoT as a Consequence of Technology R/Evolution

The IoT is a system that rely on autonomous communication of a group of physical objects. IoT is an emerging global Internet-based information architecture facilitating the exchange of services and goods. Atzori et al. [15, p. 2793] evaluated that the main domains of IoT will be: (1) Transportation and logistics domain, (2) healthcare domain, (3) smart environment (home, office and plant) domain and (4) personal and social domain. In Fig. 1 we have outlined key elements of IoT with key realms of multiverse.

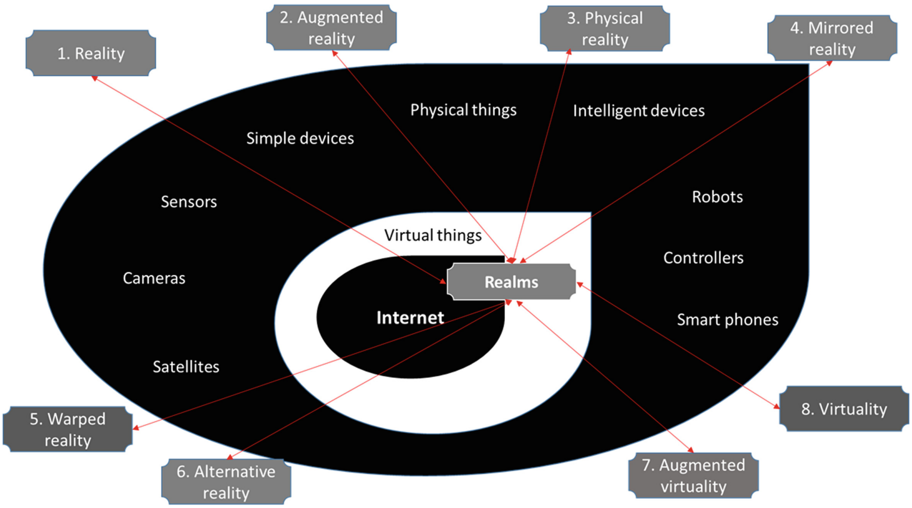


Fig. 1. Internet of Things, devices and realms of multiverse (modification from [11], p. 161]).

In Table 1 we have figured out realms of ubiquitous society. This entity is called multiversity. Table 1 tells to us that leaders, managers, planners – people responsible for running business – must understand the fundamental nature of three elements of reality: *time, space and matter* [16].

New service designs, architectures and business models are needed in the multiverse, not only in the universe. What is obvious is that managers must work in order to manage these critical eight realms of ubiquitous society.

The application are of the IoT are numerous, basically meaning smart things and smart systems such as smart homes, smart cities, smart industrial automation and smart services. IoT systems provide better productivity, efficiency and better quality to numerous service providers and industries. IoT is based on social, cultural and economic trust and associated trust management skills, which broadly speaking mean developed security services and antifragility operations. Critical issues of IoT security field are [17, p. 1505]: trusted platforms, low-complexity, encryption, access control, secure data, provenance, data confidentiality, authentication, identity management, and privacy enhancing technologies (PETs).

Table 1. Realms in the ubiquitous society and in the multiverse [15, p. 17].

Variables			Realm
1. Time	Space	Matter	Reality
2. Time	Space	No-matter	Augmented reality
3. Time	No-space	Matter	Physical reality
4. Time	No-space	No-matter	Mirrored reality
5. No-time	Space	Matter	Warped reality
6. No–time	Space	No-matter	Alternative Reality
7. No-time	No-space	Matter	Augmented Virtuality
8. No-time	No-space	No-matter	Virtuality

Security of IoT requires data confidentiality, privacy and trust. These security issues are managed by distributed intelligence, distributed systems, smart computing and communication identification systems. [17, p. 1505, p. 1508]. Finally, in Fig. 2 we have figured out the functioning pattern of markets networks and crowds. IoT can be found between these key systems of global economy. Probably there is a lot of potential for smartness between these key systems. Data, information and knowledge about communication and interaction of these systems will be vital issue for the future of management.

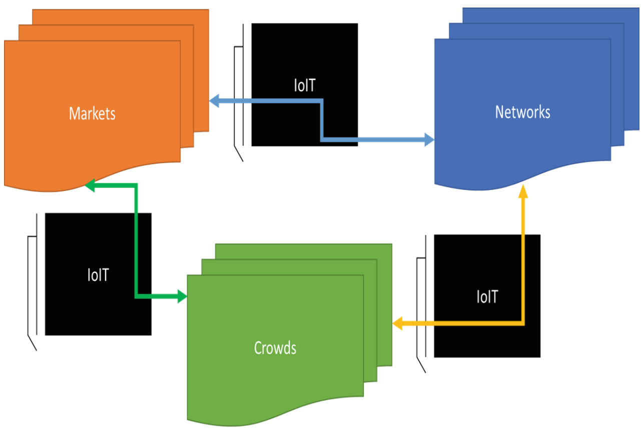


Fig. 2. The functioning pattern of markets networks and crowds.

Especially IoIt, Internet of Intelligent Things, as some experts emphasize smart Machine-to-Machine communication, provides much potential for crowdsourcing of markets and networks. IoIT provides also much potential for smart networking (between markets and networks and between various networks). We expect that one obvious consequence of IoIT will be the broader scope of deliberate democracy. Finally, the legal framework of IoT/IoIT is very vague, or it does not exist. Such issues like standardization, service design architecture, service design models, data privacy and data security create management and governance problems, which are not totally solved inside current service architectures. IoT has also become subject to power politics because of risks of cyber war, cyber terror and cyber criminality [18, p. 341, p. 347].

In Fig. 3 we present a global reference scenario for IoT-aided robotics and AI applications. We can see that IoT will be central for the collection of BigData. BigData will be collected from the (1) environment, (2) from human beings and (3) from robots and AI applications.

Figure 3 describes key elements of future management system. Robots and AI application can assist and help managers and leaders in many ways.

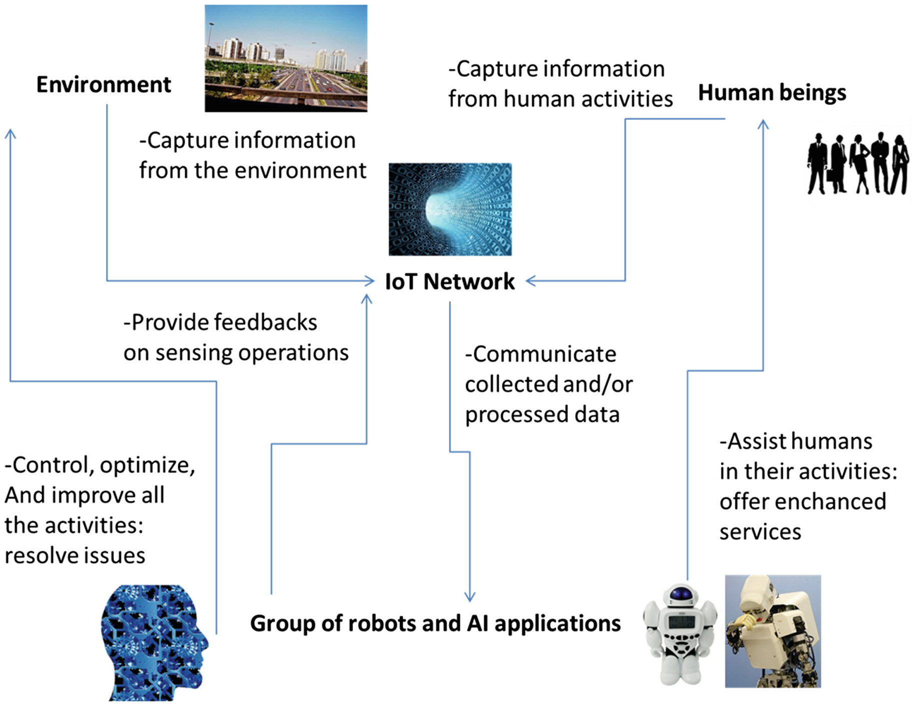


Fig. 3. A global reference scenario for IoT aided robotics and AI application (a modification of [19, p. 34]).

4 The Second Coming of Knowledge Based Decision-Making?

Seminal contributions by Simon [20] and Choo [20] and many others have showed that organizations use information and knowledge both for improving the quality of decisions and for legitimizing decisions including also those decisions made by poor knowledge. Feldman and March [22] have written one of the most persuasive articles explaining why organizations fail to use information in effective way in decision-making. According to them, organizations' knowledge behavior is rather perverse. By this they mean that although organizations "systematically gather information more information than they use, yet [they] continue to ask for more". The oversupply of information happens due to several reasons. The main reason is that organizations incentives for information are biased in sense that they tend to underestimate the costs of information gathering relative to its benefits. Typically, decisions about information are made in a different part of organization than where the actual information gathering is carried out. This division of using and gathering information enable decision-makers to launch information gathering process which may has value for them, albeit from the organizational perspective create more costs than benefits. This kind of behavior is rational for individual decision-maker as it creates an illusion of managing uncertainty. It is rational because "an intelligent decision maker knows that a decision made in the face of uncertainty will almost always be different from the choice that would have been made if the future had been precisely and accurately predicted" [22, p. 175].

Rationality of information oversupply relates also to strategic value of information. This manifests itself, for example, in cases where information is not, in the first place, used for doing sound decisions, but for persuading someone to do something. In organizational life, information is seldom neutral. Instead most information is subject to misrepresentation [Ibid, p. 176]. Worth noting is that information not only unveils some aspect of the issue at the stake, but also hides other aspects of the same issue. Feldman and March [22 p. 176] concluded that "it is better from the decision maker's point of view to have information that is not needed [in decision making] than not to have information might be needed".

Eventually, knowledge based decision-making can be seen as a widely repeated truism – a statement of obvious truth without any spesific meaning. This is because it is quite difficult to imagine what else than knowledge, could provide sound basis for organisation's decisions. Although beliefs, intuitions, and sometimes pure guesses may play important in everyday decision-making, organisations' strategic and operative choices cannot in the first place be based on them. An organization that openly admits that its' decisions are mainly pulled out of the hat does not attract trust within or outside of its borders.

Knowledge and information have probably played a critical role in organisational decision-making for as long as man has trusted on organisations, however, it was not early than the beginning of 1990 when the theory knowledge-based organization were developed. However, Grant [23] (1996) and Spender [24] (1996) laid down the cornerstone, which became known as the *knowledge-based view of the firm*. As an example of the increasing interest in knowledge as organisational resources provides the rapid growth of academic papers which used knowledge management (KM) in their theoretical lenses. In five years period just before (1990–1995) Grant's and Spender's articles, the number of papers which touched upon the knowledge management issues in peer-reviewed journals found in four data basis (Academic Search Elite, ProQuest, Elsevier Science Direct and Emerald Insight) was 87 articles, where as in five years period right after Grant's and Spender's papers (1996–2001) the number had grown to 2435.

Despite of increasing academic, as well as, practical efforts, the consensus related to knowledge in decision-making is nowhere in sight. From this paper´s view, a main divide is, whether knowledge is seen as a static asset owned by organization or as a social construction emerged from interaction. Static view on knowledge implies the managebiality of knowledge, where as social view emphasizes that knowledge cannot be managed, only enabled. Worth noting is that different approaches have different practical implications related to the role of information technology. Static view on knowledge has contributed "IT-track KM", while social view on knowledge has brought "People-track KM". "IT-track KM" treats knowledge as object that can be identified and handled in information systems. "People-track KM" deems the role of IT as useful but not critical because it emphasizes assessing, changing and improving human individual skills and/or behaviour. Related to differences in the role of IT, the two views on knowledge have also contributed two different knowledge management strategic. According to Hansen et al. [25] (1999) organisations rely on (consciously or unconsciously) either codification or personalisation knowledge management strategies. Codification strategy rests on explicit knowledge, i.e. knowledge that can be easily captured, organised and communicated [26], whereas personalisation strategy deals with tacit knowledge, i.e. knowledge that cannot be extracted from individuals [27]. Hansen et al. (1999) [25] concluded that organisations that try to exploit both strategies risk the failure of both. As an approximate division, they suggest an 80–20 split: 80 % of the organisation's knowledge practices follows one strategy, 20 % the other.

From this paper´s perspective, the most interesting question is not, however, the division of KM strategies. Instead, the identified two views on KM and the role of IT in them begs to question what possibilities come along with the emergence of Big Data. Does Big Data lay down a basis for more smart, intelligence and even wise decision-making? Does Big Data bring knowledge based decision-making into higher level?

In order to reflect the question, we need to analyse the functions of knowledge and information in decision-making. One possible useful approach to analysing decision-making is defining it as a moment which divides time into two eras, before and after decision. Broadly adapting Andersen [28], it can be argued that knowledge shapes the distinction fixed/open contingency concerning social operations (Fig. 4).

It is important to recognize that while decisions fulfill expectations they simultaneously produce insecurity in the sense that "it becomes obvious that a different decision could have been reached" [27]. To manage uncertainty related decision-making organizations' need information and knowledge to convince internal and external stakeholders that choices are made rationally. Although, conflicting interests and problems of gathering the all relevant information means that rationality in decision-making is only bounded [20, 21] Choo (1996), for example, has suggested

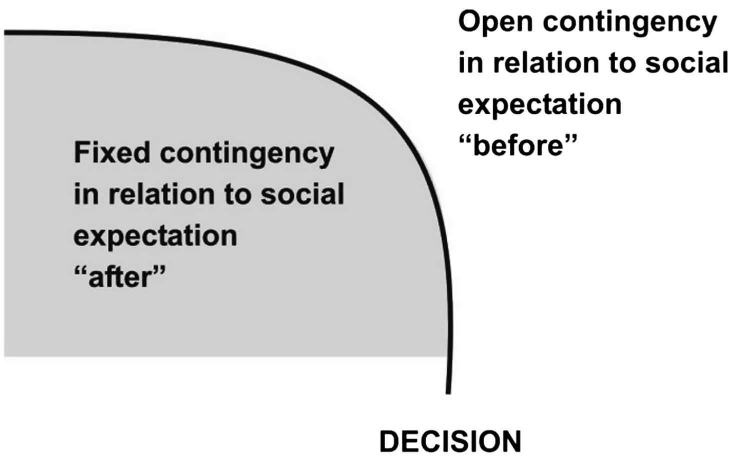


Fig. 4. Decision as a dividing system [26].

that by information and knowledge, however, it is possible to create an impression of rational and reasoned behavior, which, in turn, contributes to internal trust and to preserve external legitimacy [21, pp. 329–330]. This means that sound knowledge before decision also helps the implementation of decisions. It is also good to understand that the problem of bounded rationality is key motivation for organizational foresight activities. Brunsson [28] (1985), for example, has argued that successful management has more to do with the ability to motivate people and to create organizational culture than making rational decisions. According to Brunsson [28, p. 4] "organization´s main problem is not choosing, but it is taking organized action". Seemingly, what matters is not knowledge as "universal truth" but as "serviceable truth" [6].

The above discussion shows that information is gathered and knowledge used both for improving the quality of decisions *and* for attaining potential decision consequences. Occasionally organization's knowledge behavior is based on rationalistic ideal, whereas sometimes it is highly symbolic. Adopting the conventional view of Big Data [6], it is suggested that the true value of Big Data in decision making lies on its' ability to simultaneously promote (bounded) rational behavior (i.e. provide the best possible information) *and* to limit symbolic use of information (i.e. oversupply of information that have no value in improving decision's quality). More generally, it can be hypothesized that Big Data predicts the renaissance of knowledge management. Perhaps, the division of KM strategies into codification and personalization strategies should also be reconsidered.

5 Big Data Revolution and Smart Organizations

Next we discuss the role of Big Data with regard to organizations and start with an example. Kuper and Szymanski [30, pp. 5–6] speak about modern football as ´a numbers game´ and ground their argument because of the use of data. According to them, Opta Consulting Company was established in London in 1996 to collect match data for the English Premier League. The management consultancy´s main aim was to build a brand by creating soccer rankings. Soon The Premier League´s main sponsor paid for the so-called Opta-index and thereafter clubs and media – thus the football enthusiasts as well – got the data gathered buy OPTA for free. For instance, clubs started to learn fact they had never contemplated before: how many kilometres each player ran per match, how many tackles and passes he made, from which part of the pitch the goals were scored, and the like. The numbers revolution has been going on in football, as far as Kuper and Szymanski [29] are concerned, since twenty years now. This development has resulted to the fact, and almost unseen by fans, that the majority of the (big) clubs (at least) have arrived at statistical insights that are incrementally changing the whole nature of the game. [30 pp. 147–148, pp. 154–155].

Football clubs are organizations per se. The developments taken in using big data in the area of football indicate that you definitely need data to get ahead. If you study figures, you will see more and win more, that is. The point from football is that the beneficaries of big data are twofold: the spectators and fans on the other hand, and the clubs on the other.

This section of our paper discusses the role of big data revolution vis-á-vis organizational intelligence. Att he ouset, we argue that Big Data private and public organizations many ways. First, it can mean new business possibilities (for private business/companies) and better legitimacy and accountability(public policies and public services) at various levels. Secondly, it affects services – they can be better since the knowledge base makes it possible to access services easier or the knowledge-base can provide better focus to (co-) product services appropriately. Thirdly, it causes – if and when organizations base their actions on business intelligence - better production logic. In pratice this happens as transformation from mass-production to customized service-dominant –logic. Finally, Big Data affects organizations brand (in the case of private business/companies) and trustworthiness (in the case of public policies and public services).

As follows, we will argue that there are a number of factors affecting how the possibilities of Big Data are enhanced at organizational level. According to our view, there are number of possible drivers and possible dysfunctions that either enhance or hinder the possibilities offered by Big Data. These factors relate to the operating environment, agency, accountability, organizational coping mechanisms leadership model, information flows, innovation philosophy, production logic, and change philosophy.

As a whole, today´s organizations and their operating environments are complex entities and research-wise constantly ´on the move´. This has brought about the need to manage organizations as complex systems and to understand the logic of organizational learning organizational-wise. Given the salient nature of current economic constraints, tightening competition in business sector, problems with public sector spending and productivity, and ever growing customer demandes, the need to analyse organizational intelligence seems all the greater. In a word, organizations, dependent on the sector they operate, need to function smarter than they used to be.

Basically, the overall structure of an organization consists of leadership, strategy and foresight, people, partnerships and resources, as well as organizational processes [see e.g. 31]. This means that the modus operandi of any intelligent organisation can be defined by using these organizational features and adding the intelligent modes of action based on these elements. The conceptual idea with regard to intelligent organization needs to be clarified here. Namely, the intelligence of organization refer, to put it bluntly, to two dimensions those being knowledge management and customer-cenrted thinking throughout the organization [see e.g. 32]. This approah is somewhat different that has been put forward previously in desribing the nature of business success criteria. For intance, Peters and Waterman, Jr. [33, pp. 8–11] argued in their *Magnum Opus* that the success criteria for a successful organisation consist of various elements. These include, strategy, skills, shared values, structure, systems, style and staff (see also [34]). Currently, based on the evolvoing understanding with regard to organizations and to incorporate modern systems theory and open systems view in particular, the logic of intelligence has evolved as well and resulted in a new view to understand the role of knowledge flows in-between organizations and the role of customer needs as a foundation of service-dominant –logic. Given this, according to modern systems theory, organizations are viewed as open systems obtaining inputs from their environment, processing these inputs and producing outputs [35, pp. 39-40]; [35].

An intelligent organization is, by nature, and in essence a distributed knowledge system or sense-making community to put the idea forward by the terminology by Tsoukas [37] (2005) and Choo [38] (1998). This view holds that the resources the give organization deploys are neither given, nor discovered, but created in the process of making sense of the knowledge (e.g. [36, 40]. This comes very close to what Nonaka and Takeuchi [41] (1995) have described as a process during which tacit knowledge is converted into explicit knowledge within the structures of a given organisation. As knowledge becomes an asset in terms of organizational competitiveness, mechamisms of learning, unlearning and competence building become incalculably valuable features (e.g. [42]). This means that the traditional views of well-being at work and motivation theories with regard to work (e.g. [43, 44]) have to be re-thought and complemented with knowledge generated with regard to organisational learning and individual competencies.

Research literature indicates that performance measurement ought to be multi-dimension (e.g. [45]). Research literature also suggests that performance measurement does not necessarily mean that organizational decision-making is approproate or evidence-based (e.g. [46, pp. 6–12]): Consequently, organizations may end up in casual benchmarking, doing what seems to have worked in the past, and to follow deeply held yet uneximined ideologies. Looking from the public sector, public policy evaluation and public sector accountability point of view, the causal relation between implementation of public policies and programmes and their effects are far from self-evident (e.g. [47; 50]).

We argue that there are two kind of societal effects of the deployment of Big Data when we look at the matter from the organizations' point of view. First, there are the effects related to the objectives organizations try to achieve, i.e. services, products and manufactured goods which are their key mandates in the market. These effects can be pinpointed both to private and public sector, but from a bit different angle – namely, these effects include new business possibilities (private business/companies) and better legitimacy (public policies and public services), as well as better services for customers and service users. Secondly, there are certain effects which concern organizations themselves. Big Data enables organizations to construct their strategies on knowledge which consequently mean that they possess better foresight know-how to understand the profound changes in their operating environments. It also pave way for better production logic which incorporates the shift from mass-production to customized service-dominant –logic (e.g. [49]), which eventually means better brand and trustworthiness for the organizations as a whole. Therefore, it is noteworthy to say, that in organizational terms information – e.g. Big Data in particular – and technology are arguably one of the most important systemic changes factors, which affect organizations and organizational life. In Table 2, we have put forward nine organizational dimensions (left column) through which we we try to make sense of the possible drivers and possible dysfunctions at organizational level with regard to Big Data.

Based on Table 2, we argue that there definitely are certain organizational drivers which enhance Big Data utilization in society. As organizations operate in open system as networks, the role of information becomes truly valuable commodity. Knowledge, based on information intra-organizational information flows, and incorporated to organizational life through the mechanisms of foresight and planning, is the cornerstone of business intelligence. This calls for new understanding on the organizations′ accountability function (e.g. [50, 51]) – putting the emphasis on measuring and analysing accountability both vertically (reporting about the outputs and outcomes of an organization from bottom-up) and horizontally (reporting to customers, citizens, media, and the like). And it is important to see, that not only accountability aspects are at stake here.

This new understanding ‘requirement’ concern also innovation and change philosophy organization possesses. Innovation paradigm opens up because of the availability of information – tomorrows strategies and innovations are orchestrated ‘together’ instead of organizational siloes. We have argued earlier [3] that traditional change management models have to a certain extent come to impasse. Traditional top-down change management models do not function anymore because – to use the expression of Kets de Vries [52, p. 1] – organisations are like automobiles. They do not run themselves, except downhill. They need people to make them work. In fact, this calls for psycho-dynamic-systemic way of looking at people in organisations and a new focus on elusive micro-processes that take place in organisations. This is precisely why change management ought to be conceived two-dimensionally – it concerns individuals working within an organisation as well as the organisation which is about to change (e.g. [37, 53, 54]). As a whole, Big data also strengthens the transformation from mass-production logic towards more customized and personalized production-logic. In order to keep ‘fit’ in the tightening competion, more focus should be put on both products and services organizations are delivering.

We have indicated possible dysfunctions of the Big Data utilization in the third column in Table 2. Hierarchical thinking, vertical accountability philosophy, the non-existence of modern foresight procedures, conventional management and leadership mechanisms and skills, inter-organizational information understanding, closed, single-organization –based innovation thinking, and phase-based & linear change philosophy in organizations are examples of dysfunctions which can be detected when and if the possibilities of Big Data are not put into practice. Finally, we might add that the use of Big Data and the growing know-how about its limits strengthen organizational resilience. According to McManus et al. [55] (2008), for instance, the task of building more resilient organizations is complicated by an inability to translate the concept of resilience into tangible working constructs for organizations. In fact, resilience is often considered to be a crisis or emergency management issue and the link between creating resilient day-to-day operations and having a resilient crisis response and recovery is typically not well understood by organizations. We would like to add that resilience can be defined as an organization's capacity to anticipate disruptions, adapt to disruptive events, and create lasting value in a turbulent environment (e.g. [3]). Organizational resilience is thus the ability of an organization to overcome an internal or external shock and to return to a stable state [56]. Needless to say, resilience is the key feature of smart organization. The main point is the resilience does not occur by accident or by chance. It is the effect of smart actions and smart leadership. The capacity of resilience must be developed by smart organizational decisions.

6 Reflections on Knowledge Management in Organizations

IoT and Big Data are key drivers for change when organizations re-organize their knowledge-management practices. Tsoukas [37, pp. 110–111] emphasizes to important things in this respect. According to him, organizations are distributed knowledge systems, which means that organizations have to take a system-based view if they intend to meet the challenges of the IoT and Big Data. On the other hand, Tsoukas underlines the fact knowledge truly is a resource for organizations – it is a resource,

Table 2. Organizational dimensions as possible drivers and dysfunctions enhancing/limiting the use of Big Data.

Dimension	Possible drivers enhancing *Big Data* utilization	Possible dysfunctions limiting *Big Data* utilization
Interpretation of operating environment	Open system	Closed system
Agengy	Network Organizations as information flows	Hierarchy Single organizations
Accountability	Horizontal + vertical	Vertical
Organizational coping mechanism	Foresight-based resilience	Retrospective analysis –based rigidity
Leadership	Business intelligence	Conventional management and leadership
Information flows	Intra-organizational	Inter-organizational
Innovation philosophy	Open	Closed
Production logic	Service-dominant –logic, "customers first"	Taylorian production ideal "productivity first"
Change philosophy	Immanent, emergent, cyclical	Phase-based, linear

which is neither given, nor discovered, but actually created. This means that knowledge management practices should be developed with a system-based holistic view. The heart of this renewal of knowledge management practices lies at learning and competence-building [57, pp. 8-10]. This means that organizations need to pay attention how they compete, make decisions, apply the principles of organizational learning, connect and relate new and existing information, and finally how they monitor their success and effectiveness.

The above discussion shows that information is gathered and knowledge used both for improving the quality of decisions and for attaining potential decision consequences. Occasionally organization's knowledge behavior is based on rationalistic ideal, whereas sometimes it is highly symbolic. Adopting the conventional view of Big Data [6, 58], it is suggested that the true value of Big Data in decision making lies on its' ability to simultaneously promote (bounded) rational behavior (i.e. provide the best possible information) and to limit symbolic use of information (i.e. oversupply of information that have no value in improving decision's quality). Big Data creates value for knowledge management particularly as it provides transparency in organizational decisions (e.g. [58]). It makes information accessible across organization and therefore significantly reduces information search and processing time. At best Big Data promotes information gathering and its conversion into organizational knowledge assets. More generally, it can be hypothesized that Big Data predicts the renaissance of knowledge management. Perhaps, the division of KM strategies into codification and personalization strategies should also be reconsidered. Moreover, it is reasonable to speak of paradigmatic changes if we consider IoT and Big Data as the drivers of this paradigmatic change. In order to understand the paradigmatic changes that will take place in knowledge management practices, we would like to think of it really as a paradigm shift, even though it is always rather controversial to speak about paradigms as clear-cut epochs. The paradigm definitions are usually far from clear, precise, and determined, and of course there is a lot of room for scholarly hair-splitting and confusion here. We think that organizations are, in terms of knowledge management practices, moving from knowledge management based on information storage & decision making paradigm towards more detailed and sophisticated knowledge management practices which will be based on value-creation and business process renewal [57–59]. Whereas the older paradigm prevailed in the world of closed systems and semi-technological operating environment, the value-based knowledge utilization paradigm encapsulates the multi-dimensional aspects of the IoT and Big Data, such as the knowledge architecture, heterogeneity, scalability, look up, dynamic mashup, security and privacy, communication protocols, social networking identification, social networking management, and trustworthiness issues [60]. The evolving new knowledge management paradigm is based on the fact that the IoT and Big Data have an enormous effect on organizations´ business processes and business architectures aiming at applying value-based approaches to develop business modelling of new solutions based on the IoT in particular (see also [61]). One of the most promising approaches in this field is to connect SECI approach and Activity Theory via cascading modes of communication [62].

7 Conclusions

Bearing in mind their importance already today, IoT and Big Data most definitely are key factors affecting societal development in the future. Private and public organizations have begun to gain critical insights from the Big Data and ubiquitous technology through various management systems. Basically, the issue at stake here is the fact that it is not just the question how to manage and control the technological possibilities. The development also concern leadership functions. Namely, taking seriously Internet of Things and ubiquitous technology may lead towards the revolution of digitalization which effects on management processes in organizations. The deployment of on-going key processes call for leadership on many organizational levels. Both the utilization and the development of technologies are the key challenges in the revolution. To conclude, the key aspects of digital revolution in management process are to be considered as smart solutions in the future. Organizational processes form the base for the knowledge-based decision-making. Developing and utilizing smart solutions – like the utilization of Big Data – emphasize the importance of open system thinking. Digitalized services can for instance create new interfaces between service providers and users. Service users create social value while they are participating in co-producing activities. Hence, the IoT (or in some contexts IoIT) and Big Data undoubtedly strengthen the role of participation in service production, service economy, innovativeness in-between organizations (as a joint processes) and leadership models incorporated in service-dominant –logic.

IoT, Big Data, and especially digitalization bring about the renaissance of knowledge in decision-making. At organizational level, smart organizations do no rely on knowledge production, but focus on knowledge integration instead. Knowledge integration becomes a key part of management systems. This also means that seminal theories with regard to decision-making and knowledge management do not suffice anymore. What is needed a new understanding of organizations functioning in the framework of open systems. Open systems are interlinked with each other by boundaries constituted and manifested by knowledge. Managing these boundaries require that knowledge is exchanged, traded and made understandable in organizations (e.g. [62]). Moreover, the true value of Big Data in decision-making and in organizational terms lies on its' ability to simultaneously promote (bounded) rational behavior (i.e. provide the best possible information) *and* to limit symbolic use of information (i.e. oversupply of information that have no value in improving decision's quality). This also affects organizations´ ability for resilience. We think that resilience should be seen as an organization's capacity to anticipate disruptions, adapt to disruptive events, and create lasting value in a turbulent environment. "Built to last" and "built to be changed in modular way" are broader management issues, which should be planned carefully to develop visionary organizations. To conclude, looking from management point of view, there is a growing need to develop abilities to act in changing, not easy to forecasted and non-linear situations due to the complexity related to utilization and developing digitalization. Authentic and clinical leadership involves components such as awareness, unbiased processing, action, and relations ([63–66]). Authentic leaders are deeply aware of how they think and behave and are perceived by others as being aware of their own and others' values/moral perspectives, knowledge, and strengths, and aware of the context in which they operate [67].

References

1. Beck, U.: From industrial society to the risk society: questions of survival, social structure and ecological enlightenment. Theor. Cult. Soc. **9**(1), 97–123 (1992)
2. Cameron, K.S.: Effectiveness as paradox: consensus and conflict in conceptions of organizational effectiveness. Manag. Sci. **32**(5), 539–554 (1986)
3. Virtanen, P., Kaivo-oja, J.: public service systems and emerging systemic governance challenges. Working paper. Submitted to be published in the Journal of Service Theory and Practice (2014)
4. Boyd, D., Crawford, K.: Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. Inf. Commun. Soc. **15**(5), 279–662 (2012)
5. Zalavsky, A., Perera, C., Georgakopoulos, D.: Sensing as a service and big data. In: Proceedings of the International Conference on Advances in Cloud Computing (ACC), Bangalore, India (2013). ref: arXiv:1301.0159x
6. Mayer-Schönberger, V., Cukier, K.: Big Data: A Revolution That Will Transform How We Live, Work, and Think. Houghton Mifflin Harcourt Publishing Company, New York (2013)
7. Öztayşi, B., Baysan, S., Akpinar, F.: Radio Frequency Identification (RFID) in Hospitality. Technovation **29**(9), 618–624 (2009)
8. Weiser, M.: The computer for the 21st Century. Sci. Am. **3**, 3–11 (1991)
9. Zuehlke, D.: SmartFactory – towards a factory-of-things. Annu. Rev. Control **34**(1), 129–138 (2010)
10. Miranda, L.C.M., Lima, C.A.S.: Trends and cycles of the internet evolution and worldwide impacts. Technol. Forecast. Soc. Change **79**(4), 744–765 (2012)
11. Chen, Y., Hu, H.: Internet of intelligent things and robot as service. Simul. Model. Pract. Theory **34**, 159–171 (2013)
12. Mee, L.Y., Huei, C.T.: A profile of the internet shoppers: evidence from nine countries. Telematics Inform. **32**(2), 344–354 (2015)
13. Atzori, L., Iera, A., Morabito, G.: The internet of things: a survey. Comput. Netw. **54**(15), 2787–2805 (2010)
14. eMarketer: Retail Sales Worldwide Will Top $22 Trillion This Year Ecommerce eclipses $1.3 trillion, led by China and US (2014). Web: http://www.emarketer.com/Article/Retail-Sales-Worldwide-Will-Top-22-Trillion-This-Year/1011765#sthash.yrzMAR3s.dpuf
15. Pine II, B.J., Korn, K.C.: Infinite Possibility: Creating Customer Value on the Digital Frontier. Barrett-Koehler Publishers, San Francisco (2011)
16. Miorandi, D., Sicari, S., De Pellegrini, F., Chlamtac, I.: Internet of things: vision, applications and research challenges. Ad Hoc Netw. **10**(7), 1497–1515 (2012)
17. Weber, R.H.: Internet of things - Governace quo vadis? Comput. Law Secur. Rev. **29**(4), 341–347 (2013)
18. Grieco, L.A., Rizzo, A., Colucci, S., Sicari, S., Piro, G., Di Paola, D., Boggia, G.: IoT-aided robotics applications: technological implications, target domains and open issues. Comput. Commun. **54**(1), 32–47 (2014)
19. Simon, H.A.: Making management decisions: the role of intuition and emotion. Acad. Manag. Executive **1**(1), 57–64 (1987)
20. Choo, C.W.: The knowing organization: how organizations use information to construct meaning, create knowledge and make decisions. Int. J. Inf. Manag. **16**(5), 329–340 (1996)

21. Feldman, M.S., March, J.G.: Information in organizations as signal and symbol. Adm. Sci. Q. **26**(2), 171–186 (1981)
22. Grant, R.M.: Toward a knowledge-based theory of the firm. Strateg. Manag. J. **17**(2), 109–122 (1996)
23. Spender, J.C.: Making knowledge the basis of a dynamic theory of the firm. Strateg. Manag. J. **17**(2), 45–62 (1996)
24. Hansen, M.T., Nohria, N., Tierney, T.: What´s your strategy for managing knowledge. Harvard Bus. Rev. **77**(2), 106–116 (1999)
25. Firestone, J.M.: Key issues in knowledge management. Knowl. Innovation J. KCMI **1**(3), 8–38 (2001)
26. Nonaka, I.: A dynamic theory of organizational knowledge creation. Organ. Sci. **5**(1), 14–37 (1994)
27. Andersen, N.: The undecidability of decision. In: Bakken, T., Hernes, T. (eds.) Autopoietic Organization Theory: Drawing on Niklas Luhmann's Social Systems Perspective, pp. 235–258. Copenhagen Business School Press, Oslo (2003)
28. Brunsson, N.: The Irrational Organization: Irrationality as a Basis for Organizational Action and Change. Wiley, Chichester (1985)
29. Demarest, M.: Understanding knowledge management. Long Range Plan. **30**(3), 374–384 (1997)
30. Kuper, S., Szymanski, S.: Soccernomics. Nation Books, New York (2012)
31. Roberts, J.: The Modern Firm: Organizational Design for Performance and Growth. Oxford University Press, Oxford (2004)
32. Virtanen, P., Stenvall, J.: Älykäs Julkinen Organisaatio [In Finnish, An Intelligent Public Organisation]. Tietosanoma, Helsinki (2014)
33. Peters, T.J., Waterman Jr., R.H.: In Search of Excellence: Lessons from America´s Best-run Companies. Harper-Row Publishers, San Francisco (1982)
34. Porter, M.E.: Competitive Strategy. The Free Press, New York (1980)
35. Harrison, M.I., Shirom, A.: Organizational Diagnosis and Assessment: Bridging Theory and Practice. Sage Publications, Thousand Oaks (1999)
36. Schwaninger, M.: Intelligent organizations: an integrative framework. Syst. Res. Behav. Sci. **18**(4), 137–158 (2001)
37. Tsoukas, H.: Complex Knowledege: Studies in Organizational Epistemology. Oxford University Press, Oxford (2005)
38. Choo, C.W.: The knowing Organization: How Organizations Use Information to Construct Meaning, Create Knowledge, and Make Decisions. Oxford University Press, Oxford (1998)
39. Weick, K.E., Sutcliffe, K.E., Obstfeld, D.: Organizing and the Process of Sensemaking. Organ. Sci. **16**(4), 409–421 (2005)
40. Weick, K.E.: Sensemaking in Organisations. Sage Publications, London (1995)
41. Nonaka, I., Takeuchi, H.: The Knowledge-creating Company: How Japanese Companies Create the Dynamics of Innovation. Oxford University Press, Oxford (1995)
42. Argyris, C.: Organizational Traps: Leadership, Culture, Organizational Design. Oxford University Press, Oxford (2010)
43. Karasek, R.: Job Demand, Job Decision Latitude and Mental Strain: Implications for Job Redesign. Adm. Sci. Q. **24**(2), 285–308 (1979)
44. Oldham, G.R., Hackman, J.R.: How job characteristics theory happened? In: Smith, K.G., Hitt, M.A. (eds.) Great Minds in Management: The Process of Theory Development, pp. 151–170. Oxford University Press, Oxford (2005)
45. Kaplan, R.S., Norton, D.P.: The Balanced Scorecard: Translating Strategy into Action. Harvard Business School Press, Boston (1996)

46. Pfeffer, J., Sutton, R.I.: Hard Facts, Dangerous Half-truths and Total Nonsense: Profiting from Evidence-based Management. Harvard Business School Press, Boston (2006)
47. Vedung, E.: Public Policy and Program Evaluation. Transaction Publishers, New Brunswick (1997)
48. Bouckaert, G., Halligan, J.: Managing Performance: International Comparisons. Routledge, London (2008)
49. Vargo, S.L., Lusch, R.F.: Evolving to a new dominant logic for marketing. J. Mark. **68**(1), 1–17 (2004)
50. Schillemans, T.: Does horizontal accountability work? Evaluating potential remedies for the accountability deficit of agencies. Adm. Soc. **43**(4), 158–387 (2011)
51. Bovens, M.: Public Accountability. In: Ferlie, E., Lynn Jr., L.E., Pollitt, C. (eds.) The Oxford Handbook of Public Management, pp. 182–208. Oxford University Press, Oxford (2005)
52. Kets de Vries, M.: The Leadership Mystique: Leading Behaviour in the Human Enterprise. Prentice Hall, Harlow (2006)
53. Kets de Vries, M.: Reflections on Groups and Organizations. Jossey-Bass, San Francisco (2011)
54. Hernes, T.: A Process Theory of Organization. Oxford University Press, Oxford (2014)
55. McManus, S., Seville, E., Vargo, J., Brunsdon, D.: Facilitated process for improving organizational resilience. Nat. Hazards Rev. **9**(2), 81–90 (2008)
56. Hamel, G., Välikangas, L.: The quest for resilience. Harvard Bus. Rev. **81**(9), 52–63 (2003)
57. McKenzie, J., van Winkelen, C.: Understanding Knowledgeable Organization: Nurturing Knowledge Competence. Thomson, London (2004)
58. McKinsey & Company: Big data: The Next Frontier for Innovation, Competition, and Productivity (2011). http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation
59. Glova, J., Sabol, T., Vajda, V.: Business models for the internet of things environment. Procedia Econ. Finan. **15**, 1122–1129 (2014)
60. Mashal, I., Alsyryrah, O., Chung, T.-Y., Yang, C.-Z., Kuo, W.-H., Agrawal, D.P.: Choices for interaction with things on Internet and underlying issues. Ad Hoc Netw. **28**, 68–90 (2015)
61. Dar,K.,, Taherkordi, A., Baraki, H., Eliassen, F., Geihs, K.: A resource oriented integration architecture for the internet of things: a business process perspective. Pervasive Mobile Comput. (2015). In press
62. Wu, P.H., Uden, L.: Knowledge creation process as communication – connecting SECI and activity theory via cascading modes of communication. In: KMO 2015, pp. 403–412 (2015)
63. Krone, O., Syväjärvi, A., Stenvall, J.: Knowledge integration for enterprise resources planning application design. Knowl. Process Manag. **16**(1), 1–12 (2009)
64. Algera, P.M., Lips-Wiersma, M.: Radical authentic leadership: co-creating the conditions under which all members of the organization can be authentic. Leadersh. Q. **23**(1), 118–131 (2012)
65. Gardner, W.L., Avolio, B.J., Luthans, F., May, D.R., Walumbwa, F.: "Can you see the real me?" A self-based model of authentic leader and follower development. Leadersh. Q. **16**(3), 343–372 (2005)
66. Ladkin, D., Taylor, S.T.: Enacting the 'true self': towards a theory of embodied authentic leadership. Leadersh. Q. **21**(1), 64–74 (2010)
67. Syväjärvi, A., Uusiautti, S., Perttula, J., Stenvall, J., Määttä, K.: The reification of caring leadership in knowledge organizations research. J. Organ. Psychol. Educ. Stud. **3**(2), 93–105 (2014)

Internet of Things Backed by Knowledge Management for Smart Home

Adriana Marcela Vega Escobar(✉), Francisco Santamaria, and Edwin Rivas Trujillo

Universidad Distrital Francisco José de Caldas, Bogotá, Colombia
amvescobar@gmail.com,
{fsantamariap,erivas}@udistrital.edu.co

Abstract. This paper shows the implementation of Knowledge Management as a tool to get the best of technologies for home electric management through a domotics prototype that controls a light remotely in real time and according to user needs, which was designed and developed focused on the concept Internet of Things. An analysis of the prototype operation and its integration with Knowledge Management is performed for the user who employs it, proposing a diagram of knowledge focused on the axes: externalization, combination, internalization and socialization. In the diagram, the processes are linked to generate knowledge and facilitate attachment with information and communication technologies focused on housing. In this way, the use of smart technologies together with Knowledge Management, and the incorporation of social and cultural variables facilitates the implementation of Home Energy Management Systems.

Keywords: Knowledge management · Internet of things · Domotics · Home management systems

1 Introduction

Information Technology and Communications have allowed a better access to the internet and home devices. Because of this, it is appropriate to enter to this new digital culture by means of knowledge management to efficiently manage electric-electronic equipment, which provides more and better quality of life at the household level. In this way, the user of the equipment obtains different benefits, and easily interacts with the environment, decreasing the digital gap. In this sense, a prototype using internet of things has been designed and developed to operate a light, useful in housing because it allows improving the following aspects with respect to the traditional bulb: remote management to turn light on and off, proper use to extend its lifetime, and monitoring by user in real time. Moreover, it has other implications as potential energy savings to improve energy efficiency of the electric system by reducing the electricity bill, the maintenance needs, and the environmental impact, and it offers opportunities for innovation, among other benefits.

Moreover, the operation of such systems changes the way people think about the use of things. Therefore, it is very important to analyze their inclusion through tools such as knowledge management focusing on social and cultural variables. This helps to

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 514–527, 2015.
DOI: 10.1007/978-3-319-21009-4_39

reinforce the interest of people towards the elements of the residential environment, ensuring that they feel as active users in the process, and allowing the integration between the technical aspects and issues framed in knowledge management. The active participation of users demonstrates viability in the implementation of energy systems, since it involves changes in the social and cultural behavior of people [1].

The increment of connected devices requires an improvement in the ability to process data and the capacity to interpret them for decision-making becomes a fundamental asset for users of these systems. The integration of mental and technological spaces is important for the new energy context at the household level. Therefore, knowledge management related with internet of things is a useful strategy for effective interaction with users and it takes advantage of technology to help people, for these control their household electrical devices according to their needs but especially an analytical knowledge spurred by benefits in different aspects achieved by good energy management based on the information emitted by objects. For this reason the internet of things should be designed to adapt to the different sources of Energy supply includes, the types of loads connected in homes, allowing the traditional power grid improve energy efficiency, adding intelligence information, converting it in the internet of things energy (ITE) [2].

2 Definitions

The definitions of main terms used in the paper are presented in this section. Assertive integration of knowledge management, internet of things, automation based in domotics, devices like "*arduino*" and software, can be used to improve Home Energy Management Systems and facilitate their implementation.

Knowledge management. Knowledge management transfers knowledge of where it is generated to the place where it is used, which implies sharing it, using it, valuing it and assimilating it [3]. The sequence of knowledge management [4], refers to issues such as: Data (real event based on a set of discrete factors) [5], Information (Communication on message with emitter and receiver) [6], Knowledge (Blend of expertise, values, information and "know how") [7], Suitability (Applied, integrated and assimilated knowledge) [8] and Know (Improved suitability to develop criterion in the decision making) [8]. Nonaka and Takeuchi [9], define tacit knowledge as the property of each person and explicit knowledge as the one transmitted with formal language, meaning that knowledge is created when there is a transformation of tacit knowledge into explicit [10], to transform it again in tacit.

Internet of things. Internet of things is a tendency for the development of information technology and communications using the interconnection of physical objects with humans to provide a service. Its main pillars are: identification, communication and interaction [11]. The technology allows to connect objects with each other, which makes it barriers real and virtual worlds unite and people to interact with many devices connected by means of communication networks and real-time sensor [12].

Home Energy Management Systems. Home Energy Management Systems (HEMS) use domestic electric energy from energy efficiency and work with Advanced Metering

Infrastructure (AMI), communication, software and Smart Grid [13]. Also, they connect domestic devices for remote or centralized use based on the combination of origin network and internet as a real time energy saving source [14]. The systems use protocols for communication between devices. These may be the open standard type (free use for all), standard licensed (open to all licensed) or proprietary (exclusive use of proprietary manufacturer) [15] and are based on the Model OSI (Open System Interconnection) of communications [16].

Domotics. Domotics is the automation and control of devices and electronic systems such as lighting, air conditioning, blinds, irrigation, etc. centralized and/or remotely. It works with Advanced Metering Infrastructure and Smart Meter [17], and helps consumers to reduce energy consumption in their houses by monitoring and controlling household appliances in accordance with energy demand [18]. These appliances are integrated into a wireless control network that can be IEEE 802.15.4, Z-Wave or Bluetooth, using the global network management since it is scalable, has a comprehensive coverage and it is robust [19].

Arduino. Arduino is an electronic board designed to be connected to a network. It allows implementing a server with high level protocols. The board has a memory capacity for autonomous processing, compilers programming languages like C, and physical ports to interface with devices [20]. It works in conjunction with the board Ethernet Shield based on the Wiznet W5100 Ethernet chip and supports up to four simultaneous socket connections [21].

3 The Problem

Energy management systems based on automation concepts and using internet of things for their functioning require that people handling knowledge management take full advantage of these resources for their benefit. Changes in technology, competition and demand are altering shape significantly enhance their external resources as decision making of customers, that is why terms like co-creation (a collaborative activity in the development of services in which consumers contribute actively providing different elements that improve processes [22]), make important knowledge management environments based on internet of things. In this society of knowledge where proper management of data for good decision-making is very important, it is essential that users of these systems obtain the best benefits.

Focused systems only in the technical part for the operation not contribute greatly in obtaining benefits for users, however, the union of these with knowledge management makes its usefulness grows, incorporating social and cultural variables is important to address potential problems adjusting the Home Energy Management Systems. Is necessary to explore in the context of current technology for the Internet of Things through active participation of the people who interact with him, particularly in the social and cultural sphere, since innovation of physical objects produces changes in the human mind as natural reflex. Tackle management control of things requires the formalization of human behavior and learning, for example, cognitive psychology, neuroscience and the decision [2].

4 Proposal Development

In this paper, knowledge management and internet of things are used to achieve benefits on household facilities. The rapid development of internet of things is providing new perspectives to manage large amounts of heterogeneous data related to energy management for decision-making. On the other hand, challenges as monitoring, supervision and control of data, may be improved by using knowledge management emphasizing in the social and cultural aspects of people, since it is the union of technological elements with the interaction of persons.

In this section, aspects of knowledge management and internet of things are presented, through the design and development of a prototype managing a light controlled by the user according to its needs, which is part of a management system for domiciliary electric networks. This prototype is designed based on internet of things, and its implementation on a large scale by Home Energy Systems can offer comfort and economical and environmental benefits. However, to take better advantage of this prototype, it is necessary to propose a diagram of knowledge management so that interact simultaneously with the internet of things and implement elements such as co-creation, so that the benefits for the user be reflected in a right decision-making, allowing integration between the technical aspects and aspects related to knowledge management.

It will take into account the social and cultural variables which influence the prototype described to generate knowledge, focusing its main distinguishing features that help consolidation in the moment they implement the Home Energy Management Systems. The next sections explain the proposed diagram of knowledge management as well as the design, development and implementation of the physical prototype of the controlled light and the interaction that the knowledge management and the prototype must have for obtaining the best advantages.

4.1 Knowledge Management Diagram

The evolution of technologies, communications and computing in everyday objects of housing facilitates better interaction between them, so it is necessary for users who employ them have full knowledge of the implications of these elements linked to the Internet objects. The human behavior is influenced by various aspects such as price, problems with as environmental, comfort, confidence in utilities or commitment to change [23]. For this reason, the incidence to implement of these systems influenced by variables such as social and cultural affect the behavior and habits of people, adding more complexity to the decision-making [24].

The Internet of things influence how people assume their role within the context of housing. This depends on factors such as sex, age, social status, role in the family, profession, etc., i.e. social and cultural aspects. For example, young and professional people are more receptive to technological change, while those with more than 45 years are less receptive [25]. Energy strategies related to social and cultural variables should be designed and implemented to contribute to the welfare of people, taking into account their needs and expectations.

There are many external variables that can influence decision-making related with energy management; nevertheless, for this research social and cultural variables were selected. A diagram of knowledge management is proposed, which is based on the spiral of Nonaka and Takeuchi [26] where there are four elements or action axes: socialization, externalization, combination and internalization. These elements are fundamental to generate tacit and explicit knowledge. Figure 1 shows the approach performed by the authors of diagram knowledge management around the social and cultural variables for a physical prototype based on the internet of things, showing the stages of knowledge spiral and key actions to develop.

In each of the four axes focused understanding processes are made in the internet of objects to create individual mental models as follows:

- Externalization: Knowledge is generated and expressed linked with the internet of things making it understandable to users, so they understand the importance of this technological tool through analogies, metaphors or figurative or visual language. Based on this, explicit knowledge is produced.

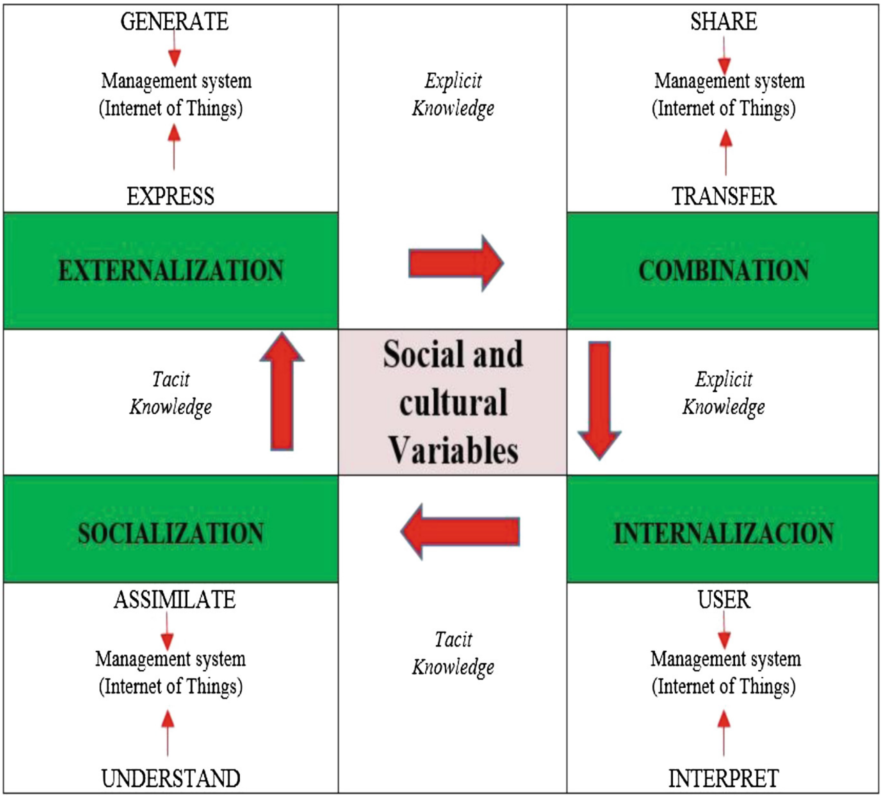


Fig. 1. Knowledge management model (source: the authors)

- Combination: Share and transfer the explicit knowledge that came from the previous step so that more people have access to this and make a better use. These are systematized and formalized concepts so that anyone can use them easily. Based on this, explicit knowledge is produced.
- Internalization: Using and interpreting the explicit knowledge that comes from the previous step to take advantage of having control over everyday objects of housing from anywhere. Experiences of others are analyzed and "learn by doing" to blend the knowledge and form mental models. Based on this, tacit knowledge is produced.
- Socialization: To assimilate and understand tacit knowledge through other elements involved in the context. Knowledge is acquired through the field of interaction by sharing experiences through imitation. Based on this, tacit knowledge is produced.

The cycle begins again at the stage of externalization, each of these iterations strengthens and make explicit and tacit knowledge of users for employing this technology and get better advantages from the environment to make feel important and the reason why be in the process of proper implementation of these systems allowing integration between the technical aspects and issues framed in knowledge management focused on the social and cultural variables. For the steps identified in Figure No. 1, to be developed in relation with social and cultural variables to generate new knowledge (explicit and tacit) and support the operation of the prototype together with knowledge management, highlighting the relevance of data for decision making, some activities are proposed in Sect. 4.3.

4.2 Physical Prototype

Domotic systems offered in the market have advantages and disadvantages; however, for every situation one or more systems are suitable for homes [27]. Likewise, research points to developments at low cost, easy to install and use matrices of sensors, motion detectors or contact switches [28]. They are based on the structure of architecture that is where "intelligence" resides of these such as architecture: centralized, decentralized, distributed and mixed or hybrid [29].

Internet of things assures the reduction of energy consumption in real time, taking advantage of data and information systems available in order to improve energy efficiency; however, besides this there are other benefits such as reducing bill payment, contribution to the environment, increasing comfort at home, etc. [30]. In the same way, proper instrumentation and smart sensors and communication protocols ensures reliable energy monitoring and control. Another important aspect to emphasize is that devices made with the conception of the Internet of Things is trust that these bring to their employers because basically have the physical layer, communication layer and presentation layer to assure system performance.

The authors designed and developed a physical prototype to control a light using Arduino. The user can remotely and centrally control the light according to its needs [31]. By means the prototype and the controlled light it is possible to visualize part of the potential of using internet of things and the importance acquired by the information technology and communications in the home environment. In Fig. 2 a picture of the

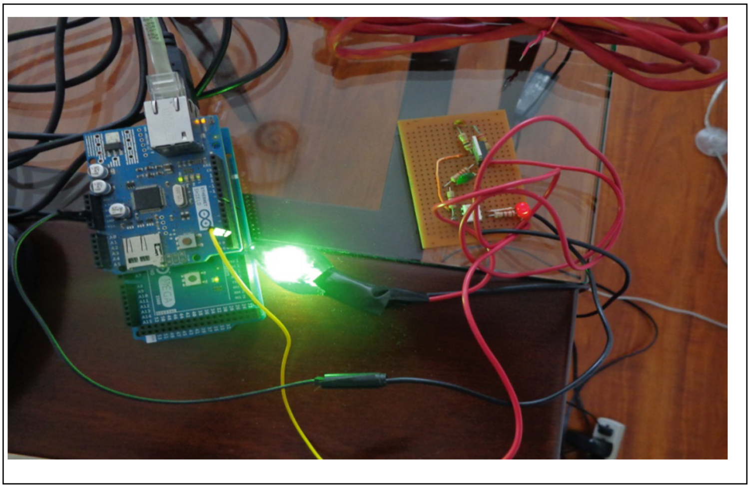


Fig. 2. Light turned on using windows azure services (source: the authors)

designed prototype is shown, which can be replicated in a full lighting system for homes, allowing the user to control and manage it in real time and get use reliable data.

Table 1 shows the components used to develop the control of a light that can be turned on and off remotely by the user anyplace in a house. It was used Arduino, because it works with simplified digital electronics, it is inexpensive, multiplatform and is open source.

Design and development of power interface. For the design of the prototype a schematic diagram of connections was used to manage alternating current with control signals and low level pin numbers were assigned, connection ports or between board Ethernet Shield, the Arduino and the light. A light of 5 W was selected, which is connected to the Arduino and the power interface can be managed in two ways: on or off [31].

In Fig. 3 it is presented the network diagram connecting the physical layer with the virtual layer, so that association of the two technologies helps to verify internet of things. A router is used as mechanism for interconnecting networks.

Software Configuration in the cloud. The virtual component or development software works with Windows Azure® server, since it is multipurpose creating in the cloud

Table 1. Elements used to manage a light.

HARDWARE	SOFTWARE
- Arduino Mega.	- Development System for Arduino.
- Connection network card Arduino Ethernet Shield	- Software Installation for Windows Azure.
- Power interface	- VisualStudio.NET 2012 Net Express.
- Device control (Light)	– Browser
– Router	
– Computer	

the respective web service. It allows for remote management of the light [31]. In the web solution called "*WindowsAzure1*", created page "*ApagarPrenderSWCloud.aspx*" The extension allows the page open application environment. The developed source code is the following:

```
using System;
using System.Collections.Generic;
using System.Linq;
using System.Web;
using System.Web.UI;
using System.Web.UI.WebControls;

namespace WebRole1
{
public partial class PrendeApagaWSCloud : System.Web.UI.Page
{tigreWSCloud.WebServiceControlArduinoCloud tigre;

protected void Button1_Click(object sender, EventArgs e)
{
tigre = new tigreWSCloud.WebServiceControlArduinoCloud();
string dir = tigre.Prende();
Response.Redirect(dir);
Button2.Enabled = true;
Button1.Enabled = false;
}

protected void Button2_Click(object sender, EventArgs e)
{
tigre = new tigreWSCloud.WebServiceControlArduinoCloud();
string dir = tigre.Apaga();
Response.Redirect(dir);
Button1.Enabled = true;
Button2.Enabled = false;
}
}
```

In the same way, the page to set the web service is called: "WebServiceControl-ArduinoCloud.asmx", the source code [31] is as follows:

```
namespace WebRole1
{
 // Description of WebServiceControlArduinoCloud
 public String Prende()
 {
 return "http://192.168.0.3:8080/?dig2=1";
 // arduino direction to turn on the lighting spot
 }
 public String Apaga()
 {
 return "http://192.168.0.3:8080/?dig2=0";
// arduino direction to turn off the lighting spot
 }
}
```

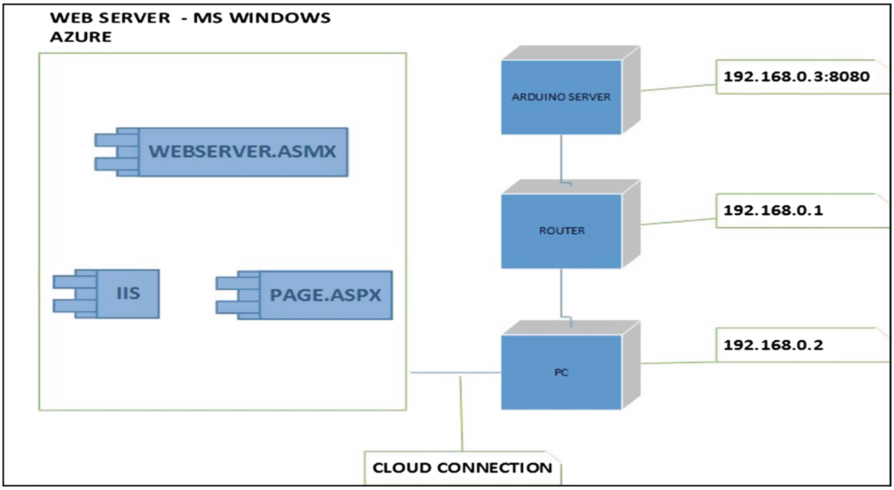


Fig. 3. Network diagram (source: the authors)

Software Configuration with Single-board microcontroller. The development of the Arduino software takes into account the information received from the website according to two states in the following way:

– State One: Send the current state of the devices plugged into the Arduino.
– State Two: Check the value of the position to turn light on or off, updating the state of the light and website. If connection to the Arduino is not established, an error message is issued.

It is important to modify the table NAT (Network Address Translation) of the router to access the Arduino from internet. In this way, requests that arrive to the public address IP (Internet Protocol) of Arduino are redirected to private IP of Arduino Ethernet Shield, to work in a coordinated way.

Operation of the prototype. To make management with internet of things in the prototype for a lighting fixture using Arduino, web service address is typed into the browser. For this case, it is: "http://adrianavega.cloudapp.net/prendeapagaWSCloud.aspx" [31]; that is the home page where warning messages, URL location, web service location, instructions or manual developed for using the prototype can be seen. At the end of the page there are buttons to remotely turn light on and off. Besides the design, development and technical operation of the prototype, it is necessary to analyze the benefits that can arise if people learn how to use these strategies at home; in this way, a proposal of the authors is presented.

4.3 Proposal for Integration of Knowledge Management

The importance of the integration between technology and knowledge management lies in the need of taking part of the user decisions. The diagram of knowledge management proposed in Sect. 4.1 facilitates these processes for people to adapt to new technologies in their homes based in social and cultural variables. A set of parameters or attributes of a society that is constantly changing is represented by the social variable, however, is necessary to identify elements related to home energy management, since this variable is emerge many others such as: economic level to acquire such systems, energy consumption habits, manners to manipulate technological devices, etc. The influence of the culture of persons it affects and modifies the use of energy management systems, for this reason the cultural variable covers aspects related to: education level using electric power, facility for making energy decisions, use of language related to energy systems, awareness of the benefits of using electricity properly.

Social and cultural adaptation are required according to each user of technology and the role they will play within the new conditions of residential electric systems and external influences that may be present. Figure 4 shows the interaction between the elements discussed in the previous sections: the prototype, the diagram of knowledge management and the user. In this way, the user interacts with the light through internet of things, but to get full advantage of the system it is necessary link with the diagram of knowledge management through activities that contribute to generate tacit and explicit knowledge, because to operate the also leads to generate energy savings, a right use of lighting, remotely manage devices at homes, etc.

Depending on characteristics of electricity management systems in homes, external variables will have direct impact on energy demand and behavior of generation. These variables bring complexity to the global model, but must be taken into account for accurately estimate forecasting models [32]. Changes in the users implementing in a

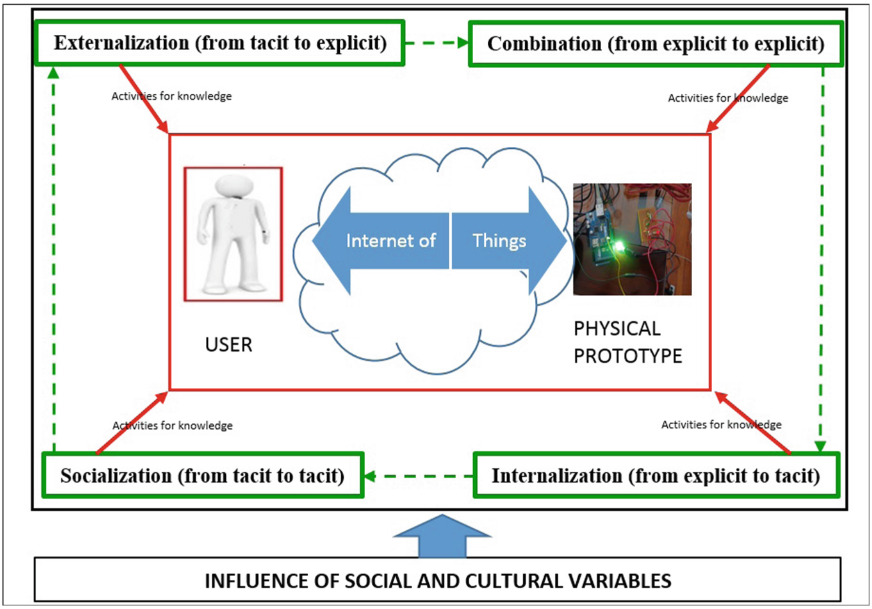


Fig. 4. Integration of knowledge management with prototype (Source: The authors)

Table 2. Actions to generate knowledge (Source: The authors)

Externalization Actions (from tacit to explicit)	***Combination Actions (from explicit to explicit)***
Transferring knowledge to an understandable form. Formally acquire a context of permanent and friendly consultation to users, including: - Promoting access to information and use of knowledge to residential users of these systems. - Let the users know the opportunities arising from implementing prototypes focused on internet of things. - A prototype for controlling a light based on internet of things, which allows the user to make decisions at home according to his needs. - Supports the understanding of all the people regardless of their social or cultural level. - Identification and modeling knowledge flows to explain sufficiently the proposed knowledge management diagram to maximize the potential of the prototype. - Brochures that can be used to teach concepts in relation with device management related to internet of things. - Web site with permanent and updated information about concepts of system operation, including virtual support, i.e. repositories, meta-information, documentation.	The knowledge is transferred to databases, so everyone can use it, regardless of its social or cultural characteristics, therefore it must be sufficiently clear and precise, from the following activities: - Divulgation in journals: Scientific and/or informative papers for anyone to understand the information about the system. - Cellars didactic information: Understandable explanations for any people, even with visual aids such as schematics and diagrams. - User manuals: Information related to the use the system. - Standards and regulations related to implement these systems. - Conducting lectures and presentations: In order to interact directly with users.
Socialization Actions (from tacit to tacit)	***Internalization Actions (from explicit to tacit)***
The experience of other people related to internet of things serves to acquire knowledge, in this phase it is proposed: - Observe how the prototype is used by other more persons skilled. - Creating email distribution lists, chats, forums, discussion groups in communities related to energy management systems. - Take into consideration the benefits of all kinds obtained by other people after implementing controls of lights through internet of things. - Know the importance of making energy decisions to have a better control over housing either remote way or direct way. - Interest to absorb the knowledge of a professional staff who knows about the issue as to manage the operation of a light based on internet of things. - Designing policies for communication and collaboration between users of the system in online forums, videos on social networks, etc.	Each person appropriates of the knowledge that he or she wants to get, the following aspects are suggested: - The user, according to his knowledge, makes consumption decisions that most suit them at different times and at any time of day, achieving different benefits. - Awareness since the early years of life of important having systems based internet of things. - Appropriation of necessary items and internet of things that allow the user to transparently interact with these systems. - People gain knowledge in an autonomous way. - Evaluations about the knowledge applied to the system. - Friendly user interface.

proper way these management systems can lead to energy savings, a more efficient use of energy, and make decisions that maximize power generation from renewable sources [33].

As a tool for adding value to the technical operations generated by the internet of things, actions are proposed to develop in the axes for generating knowledge as shown in Table 2, facilitating the integration of user with technology.

The operation of these systems produces a change in the way in which users see how these devices are used. For this reason, it is so important to analyze their inclusion through tools such as knowledge management to reinforce the value of people towards the elements of the residential environment related to internet of things. Users have to understand the reason why they must be an active part in the process of proper implementation of devices, allowing integration between the technical elements and aspects related to knowledge management. Increment the number of devices connected to the internet requires an improvement in the ability to process data, which become relevant for interpretation and decision making by the user of these devices.

5 Conclusions

The design and development of the proposed prototype confirm the advantages of internet of things, since its implementation allows the remote use of electronic devices in several locations at home. This helps people to keep continuous control of home devices, using information technologies and communication.

The prototype not only focused on turning a light on and off, but also on demonstrating that through knowledge management people acquire elements allowing them to make decisions that provide additional benefits. It is also useful at home because the user operates the system in real time and anywhere, having an overall record of objects connected, which allows controlling them.

An approach to the various modalities that can work with internet of things is done. This allows stating that a stronger culture through knowledge management and the use of technology at homes will take full advantage of electronic devices. Communication is a skill that things can be directly related to their environment through basic commands, but who manages these systems should know them to obtain the major advantage.

This type of systems based on internet of things means potential energy savings to improve energy efficiency by reducing the electricity bill, increasing lifetime, producing less impact on the environment, creating opportunities for improvement and innovation, among other benefits, but these will be achieved only if users understand the advantages through knowledge management.

In this paper only social and cultural variables were analyzed in the context of knowledge management. For a future validation, it is necessary taking into account other factors impacting home energy management systems through statistical representations; further levels of acceptance of these technologies focused on political and economic variables are interesting to analyze for the link in the value chain of the electricity sector in the new market directed towards smart grids.

Acknowledgements. The authors would like to acknowledge Research Centre and Scientific Development of the Universidad Distrital Francisco Jose de Caldas (CIDC) and GCEM Research Group, which made this study possible.

References

1. Sauter, R., Watson, J.: Strategies for the deployment of micro-generation: implications for social acceptance. Energy Policy. **35**(5), 2770–2779 (2007)
2. Ning, H., Liu, H.: Cyber-physical-social-thinking space based science and technology framework for the internet of things. Sci. China **58**, 1–19 (2015)
3. Davenport, T.H.: Saving IT's soul: human centered information management. Harvard Bus. Rev. **72**(2), 119–131 (1994)
4. Valhondo, D.: Gestión del conocimiento del mito a la realidad. Diaz de Santos, Madrid (2003)
5. Czinkota, M., Kotabe, M.: Administración de Mercadotecnia. International Thomson Editores, México (2001)
6. Chiavenato, I.: Introducción a la teoría general de la administración. McGraw-Hill, Mexico (2006)
7. Davenport, T.H., Prusak, L.: Working Knowledge: How Organizations Manage What They Know. Harvard Business School Press, Bostón (1999)
8. Beazley, H., Harden, D., Boenisch, J.S.: La Continuidad Del Conocimiento En Las Empresas. Editorial Norma, Bogotá (2003)
9. Alavi, M., Leidner, D.: Review: knowledge management and knowledge management systems: conceptual foundations and research issues. Knowl. Manage. **3**, 163–198 (2001)
10. Nonaka, I.: The knowledge creating company. Harvard Bus. Rev. **69**(6), 96–104 (1991)
11. Miorandi, D., Sicari, S., De Pellegrini, F., Chlamtac, I.: Internet of things: Vision, applications and research challenges. Ad Hoc Netw. **10**, 1497–1516 (2012)
12. Fundación de la Innovación - Bankinter, El internet de las cosas - En un mundo conectado de objetos inteligentes (2011)
13. Villa, D., Martin, C., Villanueva, F., Moya, F., Lopez, J.: A dynamically reconfigurable architecture for smart grids. IEEE Trans. Consum. Electron. **57**, 411–419 (2011)
14. Jinsoo, H., Chang-Sig, C., Lee, L.L.: More efficient home energy management system based on ZigBee communication and infrared remote controls. IEEE Trans. Consum. Electron. **57**, 85–89 (2011)
15. The Cambridge, The Cambridge Dictionary of Statistics., ISBN 0-521-81099-X. (2012)
16. Tolosa, G.: Protocolos y Modelo OSI. (March 2014). http://www.tyr.unlu.edu.ar/TYR-publica/02-Protocolos-y-OSI.pdf. (Last login: 15 November 2014)
17. Snyder, A., Gunther, E., Griffin, S.: The smart grid homeowner: an IT guru?. In: Future of Instrumentation International Workshop *(*FIIW*)*, pp. 1 – 4 (2012)
18. De Silva, L., Morikama, C., Petra, I.: State of art of smart homes. Engineering Applications of Artificial Intelligence. **25**, 1313–1321 (2012)
19. Nowak, S., Schaefer, F., Brzozowski, M., Kraemer, R., Kays, R.: Towards a convergent digital home network infrastructure. IEEE Trans. Consum. Electron. **57**, 1695–1703 (2011)
20. Arduino: Arduino, (26 Noviembre 2014). http://arduino.cc/es/Main/ArduinoEthernetShield. (Last login: 12 Dicember 2014)
21. Schmidt, M.: Arduino: A Quick Start Guide. Elsevier Pragmatic Bookshelf, North-Holland (2011)

22. López-Sánchez, J.Á., González-Mieres, C., Santos-Vijande, L.: Service innovation and custormer: co-creation: effects on performance, *Revista española de investigación de marketing,* pp. 79–101 (2013)
23. Sachdeva, A., Wallis, P.: Our demand: reducing electricity use in victoria through demans management, Monash Sustainability Institute, Melbourne (2010)
24. Zaeri, M., Sharda, N., Zahedi, A.: A five layer model for simulating a virtual power plant, de IEEE INNOVATIVE SMART GRID TECHNOLOGIES, Kuala Lumpur (2014)
25. Lorente, S.: La casa inteligente hacia un hogar interactivo y automático. Fundesco, Madrid (1991)
26. Nonaka, I., Takeuchi, H.: The Knowledge-Creating Company. Oxford University Press, New York (1995)
27. Jarman, P., Hooton, R., Walker, L., Wang, Z.: Transformer life prediction using data from units removed from service and thermal modelling (2010)
28. Berenguer, M., Giordani, M., Giraud, F., Noury, N.: Automatic detection of activities of daily living from detecting and classifying electrical events on the residential power line. In: 10th International ConferenceE-health Networking, Applications and Services, pp. 29 – 32 (2008)
29. Moore, R.: Ethernet For IEC61850. Pac World, vol. 13 (2010)
30. Shrouf, F., Miragliotta, G.: Energy management based on internet of things practices and framework for adoption in production management. Cleaner Production **100**, 1–28 (2015)
31. Vega-Escobar, A.M., Santamaría, F., Rivas-Trujillo, E.: Internet de los objetos empleando Arduino para la gestión eléctrica domiciliaria. Revista de la EAN - Escuela de Administración de Negocios, no. 77, pp. 22–42 (2014)
32. Hernandez, L.L., Baladron, C., Aguiar, J.M., Carro, B., Sanchez, A.J., Lloret, J., Massana, J.: Survey on electric power demand forecasting: future trends in smart grids, microgrids and smart buildings. IEEE Commun. Surv. Tutorials **16**(3), 1460–1495 (2014)
33. Dahlbom, B., Greer, H., Egmond, C., Jonkers, R.: Changing Energy Behaviour. Intelligent Energy Europe, Madrid (2009)

WSN Lifetime Management with the Predictive Energy Management Mechanism for the Autonomous Cooperative Smart Logistics System - A Real World Knowledge Representation Scenario

Muthuraman Thangaraj(✉) and S. Anuradha

Department of Computer Science, Madurai Kamaraj University, Madurai, India
thangarajmku@yahoo.com

Abstract. The Logistics monitoring system is one of the boons of technology innovations and cater to the fields of freight management, fleet management, workforce management and trip automation. The limited power supply of the batteries is the key concern in Wireless Sensor Network even with the alternate energy sources. The energy dissipation model of all the sensors are not the same as some of the routes have more traffic and some of the nodes play the vital role of cluster head. It is necessary to manage the energy of nodes and all the operations are to be energy-aware, to extend the lifetime of the Network. This paper discusses about a lifetime model with the energy dissipation method to predict the life of nodes and tune the algorithms accordingly, so that the entire Logistic system can be autonomous and self managed. The knowledge representation and application of knowledge both are equally important, the trend on energy-dissipation-knowledge and the trend on network-communication operation-knowledge, based on the mathematical model of the network, energy dissipation and prediction of the lifetime of the sensors in logistics domain are considered in this paper.

Keywords: Wireless sensor network · Lifetime · Energy dissipation · Predictive mechanism · Logistics monitoring

1 Introduction

Road transportation is the common local transportation mode, usually adopted everywhere. Not only local, but also across border shipping trade earns the major revenue for the country. The transportation happens in multi levels, starting from air and express delivery, freight rail, maritime (Sea Pots, Terminal) and trucking. In trucking alone, as per the recent statistics, USA earns 700 billion dollars and 9.4 billion tons of freight transported during a year [1]. The rail cargos carry mostly the raw materials across the country. The shipment cut off time depends on many factors like production delay and customer need.

The challenge here is how to represent the knowledge of logistics industry relevant to different Internet Of Things (IOT) objects with mobility, heterogeneity; energy

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 528–549, 2015.
DOI: 10.1007/978-3-319-21009-4_40

dissipation models so that the lifetime of the objects can be predicted for the consistent continuous coverage.

This paper is organized into the following structure

a. Details of the various types of sensors used in logistics.
b. Definition and discussion on the different aspects of the new vertical of autonomous cooperative logistics.
c. Realize the different avenues of smart logistics.
d. Substantiate the need for accurate knowledge representation for the smart autonomous logistics, especially the knowledge about the energy level of the smart devices.
e. Elucidate the related work in this field – abridged literature survey.
f. Explain our contribution of energy management framework AATRAL for WSN with architecture and design.
g. Propose the Energy dissipation model for the intelligent devices.
h. Put forward the predictive mechanism of lifetime management of IOT.
i. Outline the different outputs and results.
j. Summarize the summary of findings with the real world implications.

2 Sensors in Logistics

2.1 Different Types of Sensors Used in Logistics

Sensors are the tiny electronic motes to sense the environmental factors like temperature, humidity and pressure. Wireless Sensor Network (WSN) is the network of sensors connected logically without wires to communicate with each other and to propagate the message to the base station [3]. There are different types of sensors that support the logistic verticals. This section discusses on that.

A. RFID Sensors. RFID is used for identifying and tracking the tags with the goods and to transfer the identified unique identifier data wirelessly using electromagnetic fields. RFID readers are commonly used in postal cargo and libraries to track the stock in place and in workforce management as the swipe cards for the time tracking and reporting. The RFID, not only reads the barcode for loading up and unloading the goods, but also plays the role of temperature control system, which notifies the master control system about the difference in temperature, or one step further, queues the compressor to manage the temperature of the transportation of the food items during their transport with the conditioned temperature (Fig. 1).

In most of the countries every document or the letter is tagged with RFID for the delivery and supply chain management to track their whereabouts. Besides the asset management, RFID tags are used as fuel trackers and getting placed in the fuel tanks and tracks the amount of fuel and the transaction in the system. Some of the RFID tags are used in controlling the CO2 emission and helps in green transportation. RFID tags are used for innovative route planning for ensuring good customer service [4].

B. Image Sensors. The imaging sensors shoot the image of the packing to get verified at the unpacking end to make sure of the security. To scan the shipment sizes, to control

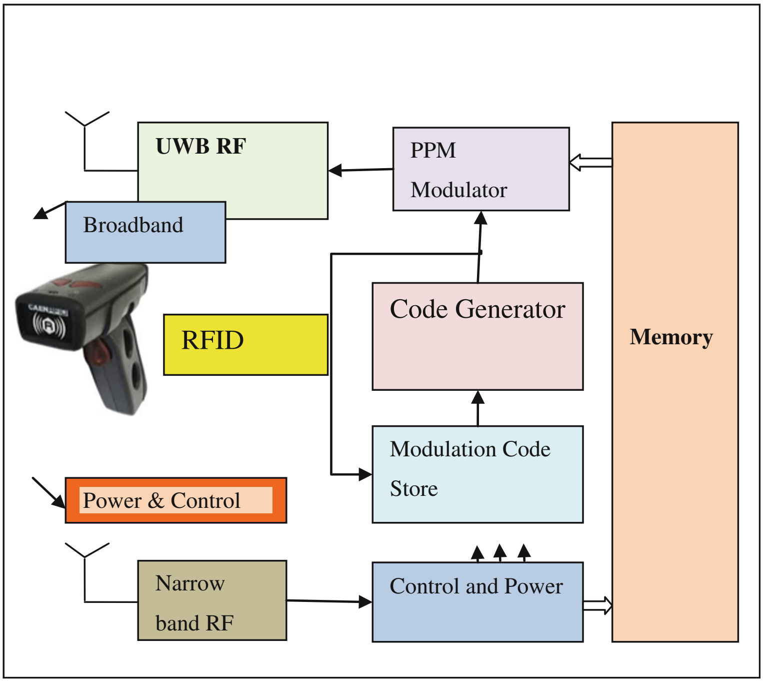


Fig. 1. Typical RFID structure.

of storing systems and to determine the level of container cargoes, the range of sensing the image, the resolution of the image, the frequency of image capturing and propagation are some of the quality parameters considered [5].

C. GPS and GPRS. With GPS (Global Positioning Sensors)/GPRS (General Packet Radio Services) in smart Logistics, the service users, service providers can track the current position of the asset at any point of time for the delivery, route planning and to ascertain the current status of the product (Fig. 2).

With the help of GPRS system, the customers can directly track their package and the service providers can easily make the trip sheet without any data entry.

D. Gas and Fuel Sensors. When the food articles, oils consignments are transported, the leakages and the gas formed out of the food are to be sensed and communicated for ensuring safety and health. Fuel sensors in the fuel tanks have multiple purposes such as measuring the fuel type, the state in which what type of fuel filled and where it is consumed and all the details [6]. Later we can see how it reduces the pain of making the fuel tax filing with every territory.

E. NFC (Near Field Communication)/NFMC (Near Field Magnetic Communication) and Smart Logistics. In logistical value chain, the NFC reader enabled phones and wireless communication play a vital role. NFC is enabled in various mobile phones and the environment will be thus set up to run the multi business applications.

Fig. 2. Typical logistics with GPS tracking

3 Autonomous Cooperative Logistics

The interdisciplinary business development process includes the supply chain management, procurement, logistics, seller and buyer markets. These diverse disciplines are integrated through the smart devices and the cloud enabled centralized applications. The cloud enabled applications with the smart devices, coordinate the legacy system in a secured and authentic way [7]. This paradigm shift need the new processes, policies, strategies and simplified solutions. It is not only the coordination between multiple systems, but also the transportation planning & pricing, optimal network flow.

The complete life cycle of autonomous cooperative logistics with the interaction of the intelligent devices and sensors is explained in Fig. 3. When the logistics items are getting stored in the warehouse, it is just scanned for the smart tag. While uploading the goods from the storehouse, it is just scanned with the RFID or simple Android phone with NFC code. Anyone can find the current stock level in the storehouse and the availability in the container from the simple mobile report.

With the cloud tag of the goods scanned, we can track the whereabouts of the logistics item, even though the goods are getting booked by different customers from different regions. The order processing system and booking system data are to be synchronized with cloud tag data so that they can categorize as per the region of distribution. Then, with the help of optimum routing mechanism at the delivery end the logistic items are classified.

The operations are based on the cloud data on the load and it is available for all the users through any of the configured devices. At any point of time the cost of the products, possession and location of the products can be visually monitored, tracked and dispatched as per the customer priorities. Here, the Ineternet of Things (IOT) with the Bigdata platforms of the intelligent products with the infrastructure of the cloud and the batch process enabled is the key technology.

The other important aspect is the integration of the autonomous cooperative Logistics with the other legacy systems of fleet management, order management, warehouse management, transportation management, reverse logistics and inventory

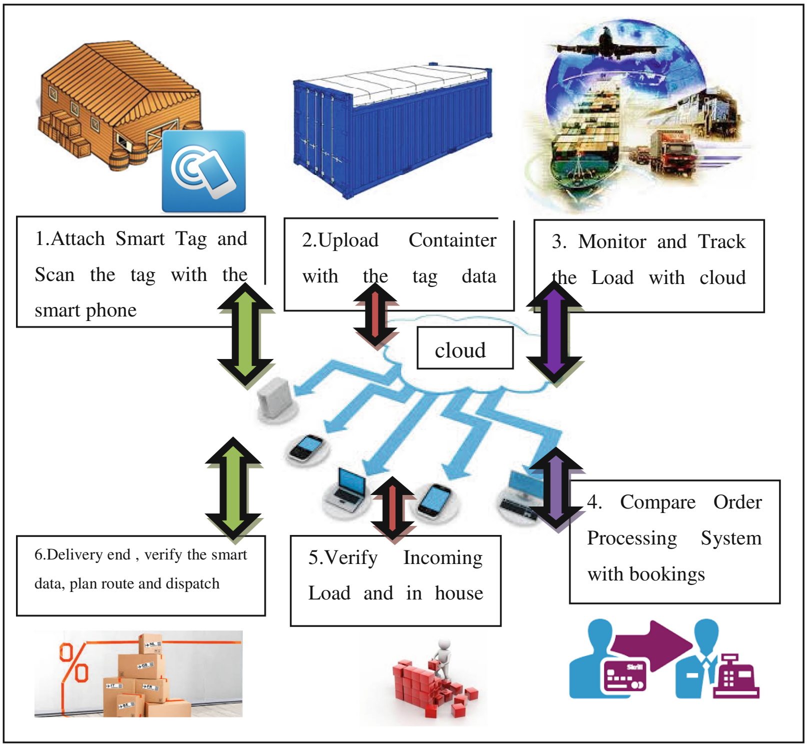


Fig. 3. Autonomous cooperative logistics

control. Autonomous control has multiple advantages over the centralized control as the decision making is distributed to local entities [6, 7]. The major limitation of the autonomous systems is the limitation of the power in the smart intelligent devices. So, the monitoring of the power at all the stages is of great importance and decisions are to be made based on that.

4 Smart Logistics

a. Asset Management. The transportation companies have their own vehicles and have tied up with the service providers. In all the cases, the fleets are to be maintained with the registration, insurance, regular maintenance and tax filing to use it. The next step is to assign the drivers for the vehicles and allotting the trips with the goods loading and unloading [8]. The trips are to be planned with the shortest path and the priority of the goods delivery commitments. To automate the process, the smart sensors are used. Auto reminder mails are designed to simplify the work.

b. Fuel Tax Filing. The fuel taxes (IFTA) in the USA are getting calculated based on the fuel purchased, the purchase place and the fuel usage location. It also depends on the state authority squatted rate, which gets released every quarter from every state.

There are some regions which are declared as non taxable regions and there are some fuels, which get concession in taxation for a particular period. The fuel type along with its Geo locations are to be tracked with the smart sensors for the maximum effectiveness [9].

c. Operation Visualization. The operation, visualization is not only query based but also include the automated alerts, workflows, new scenario case modeling, auto generated reports, and escalations on auto sensed unusual incidents. Here the workflow and the communication channel decide the operation cases [10]. Serving the clients within the committed service time is made simpler with the autonomous logistics applications of driver allocation, trip sheet and cost sheets, based on the service level agreement.

From the Fig. 4, one can perceive how in the intelligent devices, the smart sensor data is flowing into multiple legacy systems and the data is getting unified in the smart logistics. The fleet management system helps the vehicle wise trip booking, driver allocation and cost sheets. The legacy systems like Enterprise Resource Planning (ERP), Human Resource systems are to be integrated with the sales processing system for the drivers timesheet and reports of the sales pipeline and order mapping. These are together in turn to be integrated with the smart logistics of the holistic autonomous cooperative logistics.

The integration of different existing systems with the smart logistics using the unified data model of sensor data leads to the complete operation and visualization. These consist of the whole range of sales pipeline, forecast of order booking, trip sheet, reports, tracking, delivery and the actual sales. The historical data helps us to derive the

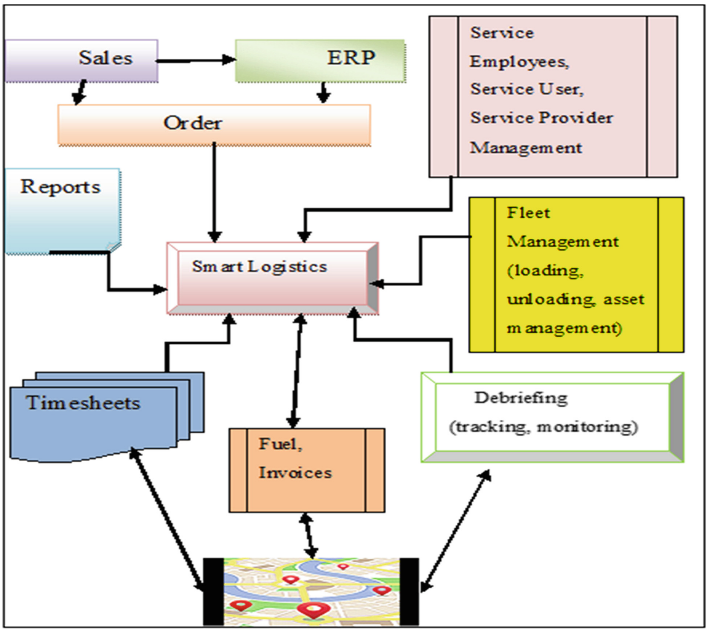


Fig. 4. Smart logistics with other external applications

trend of the sales involved with logistics from driver, vehicle, route and the needed information from the customer angle.

5 Need for Accurate Knowledge Representation for Smart Autonomous Logistics

There are transactions and interactional data which are going to mean a lot in the system. The trip sheet is going to be the base of the driver settlement which in turn has deep impact on the driver workforce service provider system. When the customer querying about the where about of the product, the interaction data stored indicate the GPRS location along with RFID tag value. Energy relevant data for all the IOT devices with their current residual energy details to decide on the usage route and replacement of the device batteries are also to be recorded. Here not only the simple data from sensors will help us, but the information, knowledge and wisdom out of it are going to support the battery replacement decision making [12]. So, the details of patterns, models, predictions, dissipation rate, average shipments, average unsuccessful deliveries, customer priorities, data on most traveled roads, the best service provider of work forces are to be made easily available for the business intelligence reports and decision.

The autonomous cooperative smart logistic system implementation with the feasibility of integration with the legacy system integration irrespective of the current technology choices, needs the basic technological entities described in Fig. 5.

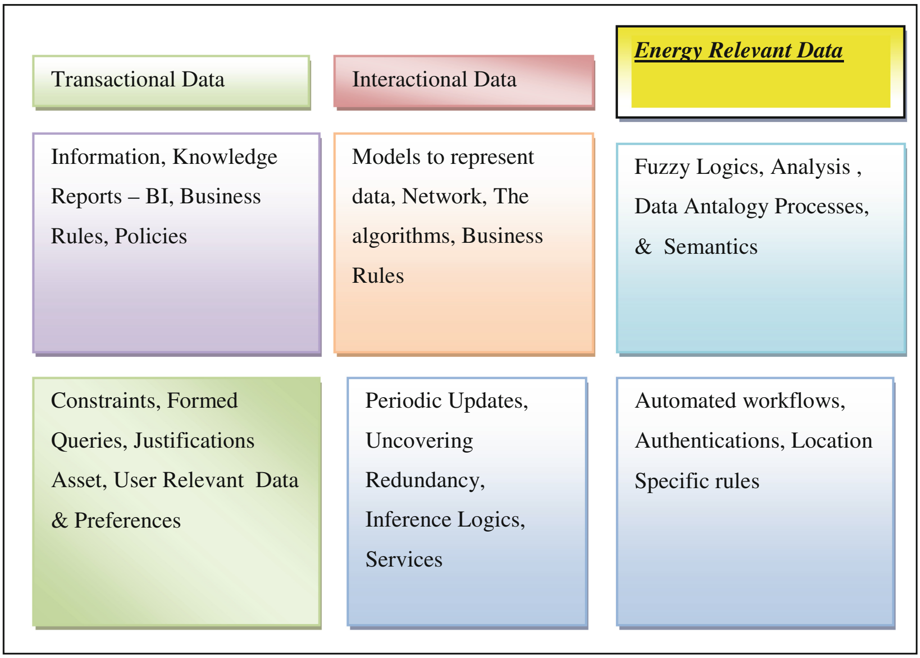


Fig. 5. Knowledge representation avenues

Automated Workflow with the work items are to be identified first which can be implemented in the WF framework of .Net or JBPM is the technology choice. Figure 7 highlights the energy relevant data of the sensors for the connectivity and coverage. The only volatile data in the smart logistics system is the data that is getting propagated by the sensors in the network. Now we seem to worry about the data about the sensors for the connectivity and coverage rather than about the data that is getting propagated with the sensors [13]. One of the critical data associated with the sensors is the energy level of the sensors. Until the **"Potential Energy Scan"** for sensors happens in the system and the **"Energy Map"** of he sensors are listed out, we may not be in a position to keep the system alive with connectivity because sensors functions with limited battery life. The energy consumption for propagation is drastically higher than the energy of sensing. So, the power level depends on the distance between the source and destination or the nearest cluster of its own type node.

Figure 6 highlights how the simple data in the system get transformed to valuable inferences, analytical data, reports, alert and mails on decision making and forecasting the business operations.

6 Related Work

a. Mathematical Modeling of Wireless Sensor Network. There are theoretical models or mathematical model of visualizing the Wireless Sensor Network (WSN) as

(a) Geometry and algebra to represent the spaces with the spatial correlation of moving nodes with the intersection of hyper planes. Varied plane shapes are considered for the WSN network plane representation with the multiple possible surfaces [13, 14].
(b) Represent the nodes as a graph of queuing nodes with random connections [14].
(c) Gaussian or Bayesian network representation with the trust estimation [15].
(d) The state of nodes may vary as sleep, active, request and response. The node modes (states) of operation are represented as Marckov chain where the random

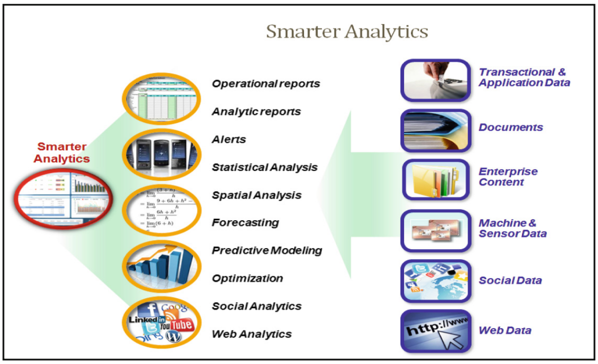


Fig. 6. Smarter analytics metaphors

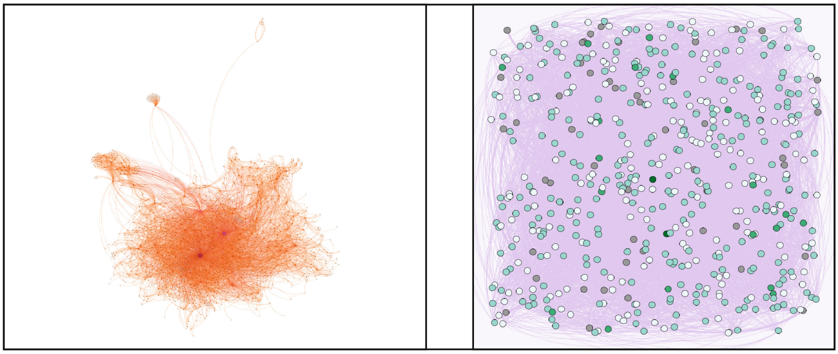


Fig. 7. graphical model of wireless sensor network with different types of nodes (base node, cluster head, sink)

variables indicate the nodes current state or mode. In all the cases, the probability factor is considered for the node to become the cluster head and to become the node to propagate the data [15].

b. Energy Dissipation Estimation. Radio model, which dynamically determines the power level setting should be used to transmit between two nodes. Using the power level setting, the network lifetime and the mean energy of the network are calculated, based on the chip specifications to ensure an accurate estimation of lifetime. Transmission can only occur at predefined discrete power levels of the amplifier. Energy dissipation for transmission is constant for a particular power level. By proposing the minimum cluster size (node degree), the minimum energy dissipation is assured in some of the works. The near energy utilization to prolong the lifetime of WSN is the focus of most of the energy dissipation initiatives considered abovet [15, 16].

c. Predictive Mechanism of Life Time Management. Most of the connectivity-aware, self healing predictive mechanisms are based on the probability model of getting the node the chance of propagation or the random factor of the node become the cluster head [17]. Rao and Biswas (2005) proposed a biologically inspired mobility model for balancing the energy overhead related to communication. The algorithm adopts a preventive approach to the creation of coverage holes due to node energy exhaustion [18].

7 Energy Management Framework Aatral

'Aatral' is the energy management framework that has been developed and tested in the lab of Madurai Kamaraj University for the wireless sensor network. It has 3 units,

1. Energy Auditing Unit
2. Energy Optimization Unit
3. Energy Harvesting Unit

The framework facilitates configuring the WSN profile with the devices and helps in monitoring and auditing the energy levels at unit level, activity level and protocol level. It allows to store them along with the time stamp so that it can be compared and a study of energy with a different possible WSN profile can be made possible. The auditing system recommends the various possible "Energy Engineering Fixes" by comparing with a benchmark. By user selection of the fix option, all the energy corrections will be implemented. Energy harvesting system facilitates the sensors to get charged with the external energy sources (Figs. 8, 9).

Fig. 8. Energy management framework aatral for WSN

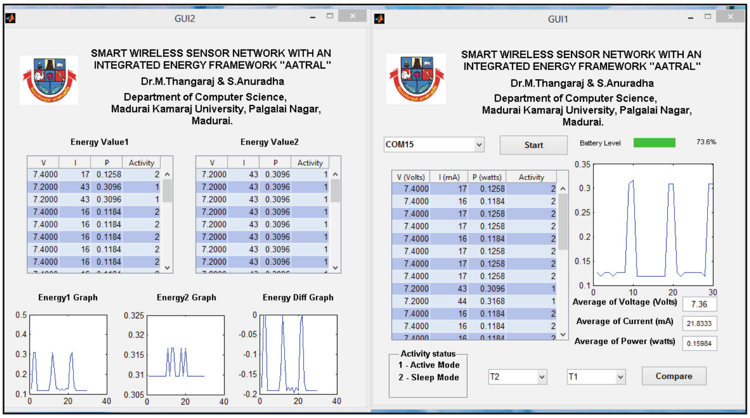


Fig. 9. Energy recordings & energy comparisons in aatral

It is an independent framework which can be plugged into any other application of wireless or smart devices to measure and optimize the energy [11]. 'Aatral' has the guided setup which makes it portable to any Smart Sensor network. Then Aatral starts managing the energy associated with that network.

8 Energy Dissipation Model of Intelligent Devices

The energy model is the one that is formed out of the factors that are used to measure the energy in the wireless sensor network system. The energy model is more accurate if it considers all the factors of protocols, device constraints, routing algorithms, cluster formation or data aggregation techniques. It again varies based on the latency delay, throughput, transmission rate, queue length, packet size etc (Table 1).

Very simple higher level equation of node energy usage can be estimated as follows

$$\begin{aligned}\text{Energy Consumption of Node} = {} & \text{Node Level Energy} \\ & + \text{Energy for Communication}\end{aligned} \tag{1}$$

Table 1. Energy dissipation factors in WSN

Energy decipiation factors	
Sensing energy	S_E
Processing energy	P_E
Propagation energy (read, propagate)	$R_E + W_E$
Peripherals energy	PH_E
Flash memory operation energy	FM_E
Routing energy (route table management, routing algorithm)	R_E
Data aggregation energy	A_E
Cluster formation and localization	C_E
Collision and Retransmission Energy	CR_E
Encryption – decryption energy	EN_E
Initial energy at node	I_E
Number of nodes and particular node	N - N_I
Frequency of transmission	F
Packet size	P
Buffer size or queue length	B or Q
Time interval of sensing	TI
Throughput of the network	T
Average delay of packets for node I	AD_I
Probability of packet loss for node I	PL_I
The probability of successful receipt of packet for node I	PS_I
Processing time of packet at node I	PT_I
Noise in data packets of node I	NO_I

$$\text{Node Level Energy} = S_E + FM_E + P_E + (R_E + W_E) + PH_E \tag{2}$$

$$\text{Energy for Communication} = R_E + A_E + CR_E + EN_E \tag{3}$$

Even though, the formulae look simple, there are other factors to be considered for the accuracy. All the nodes don't usually get the possibility of becoming the cluster head [18]. Usually in the random deployment, the nodes which receive the first hand signal information, become the cluster head. Then they do send the signals for getting connected with the next level nodes. Later corrections are introduced to rotate the sink or the cluster head to avoid more energy dissipation at one cluster node.

The collision, retransmission, the buffer size or the waiting queue size determines the successful transmission of packets. The routing algorithm chosen and the aggregation scheme chosen determines the energy on routing [19]. Even though, the energy model seems to be relevant to a node, they are more specific to a network profile than the node (Fig. 10).

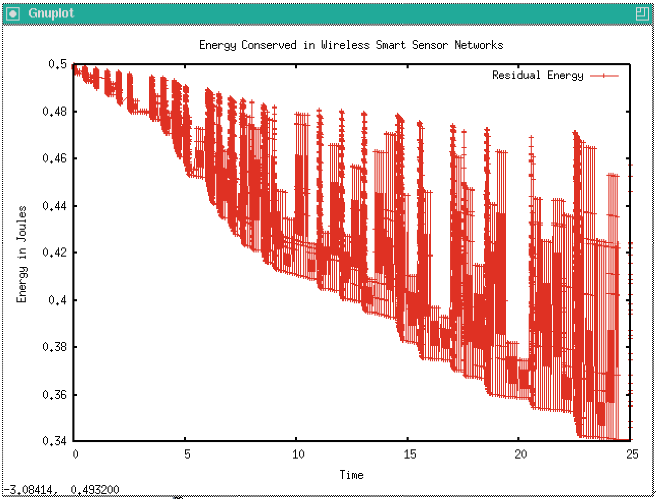


Fig. 10. Residual energy of the nodes in the WSN (Traditional Recording)

Based on the energy dissipation rate and the average energy on the nodes, the pattern or trend of energy dissipation is calculated for the system. For example, the sensors at the warehouse, because of the massive upload/download, dissipates energy more with the excess sensing and propagation, where as the tracking system works on the query based investigation, less usage or no usage of sensors when here is no query.

The following quality of service parameters, the network node population, its communication links, average hop of transmission, latency (delay in communication), and delay in the network, the throughput and baud rate (The transmission bytes for the time interval), average lifetime all are validated before and after optimization of the

WSN for fairness. The node which is far away from the base station dissipates more energy as it involves long range of communication.

$$\text{NodeLifetime} = \text{Initial Battery Capacity} \,/\, (\text{Average Current per Day}) \ast 365 \ast 24 \tag{4}$$

In the battery of the sensor, after 50 % of the threshold the energy dissipation rate goes on at a more rapid rate and not in the usual rate. Aatral has the inbuilt mechanism to find and inform the energy map of the sensors to the users.

9 Predictive Mechanism for the Life Time Management of IOT – Our Contribution

The prediction could help in selecting the cluster head, routing path, aggregation mechanism, scheduling, duty cycle based on the current energy level, state changes of the nodes. Such a prior energy level prediction ensures that the coverage won't be lost or a subnetwork totally becomes unreachable [20]. After every 3−4 duty cycles, the prediction of energy is calculated with the trend and the available data set of energy. The graphical reports of "Aatral" helps in deciding the trends and patterns of energy dissipations. The energy prediction logics are based on these patterns for the profile and the business specific rules (Figs. 11, 12).

The above pseudocode explains how the realistic energy model is formed with trend factor in the Aatral Framework

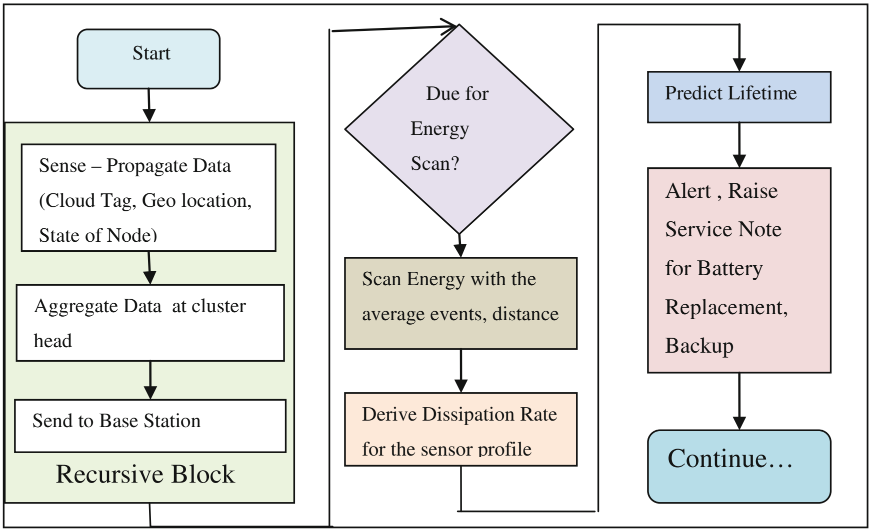


Fig. 11. Predictive life time management workflow

```
% Author                   : S. Anuradha
% Reviewed and Approved    : M. Thangaraj
% Date                     : 10-5-2014
% Copyright Reserved       : Madurai Kamaraj University
% Type of Code Segment     : PSEDUO CODE
%  Purpose                 : Aatral - Energy Framework
%                            Application on Smart Logistics
%  Data Model Refered      : Energy, WSN Profile, Node Activity, Node Communication,
%                            TRENDFACTOR, Node Energy Disipate, Node Lifetime,
%                             Customer SLA %

Begin(AatralLogistics)
Register Customer with SLA
Define WSN Profile with No. of Nodes, Protocol, Node Hardwares, Frequency,
       Modulation Scheme, Baud Rate
Define Customer Workflow as per SLA
Initialize  Node Energy Disipate, Node Communication, TRENDFACTOR, Node
           Node Lifetime, Period of Energy Scan, Action on Poor EnergyLevel

Begin(Customer WorkFlow)
    Sense_Propagate_Data(Cloud Tag, Geo location, State of Node)
    Aggregate Data  at cluster head
    Send to Base Station
If (Due For Energy Scan as per Customer SLA)
        Do PotentialEnergyScan(average events, average comm. distance,
        energy level)
        Calculate TRENDFACTOR
        Calculate NodeEnergyDisipate(average events, average comm. distance,
        energy level, TRENDFACTOR)
        Find PredictLifeTime
        Generate AutomatedReport
        Send Report to DistributionList as per SLA
        Escalate LowEnergyNodes Below Threshold
        Do Backup of Energy Records
    End
    If (Purchase Order Automation Configured in SLA)
        Place Purchase Order for Batteries
        Raise Service Request for Battery Replacement
```

Fig. 12. Predictive life time management PseudoCode

10 Results and Discussions

a. Energy Map of Sensor With Energy Level. At any point of time, Aatral helps to monitor the consolidated current energy of different sensors in the network, which helps the dashboard viewer to decide which of the sensors batteries are to be replaced or switched to harvested energy consumption mode (Table 2).

The threshold battery level is decided by the system administrator as 20 %, so the trend line in the graph indicates a line of threshold level of battery. Any of the nodes energy level is below 20 %, the node is at the risk of getting disconnected and lose the network coverage (Fig. 13).

b. Number of Events per Sensor Tracking. On an average number of events that are happening for a sensor in the network really matter because it determines the energy dissipation trend.

It is obvious from the Fig. 14, the sensors RF1-Store, RF2-Warehouse, RF3-Warehouse, IM2-Store are more loaded with higher possibility of wearing out soon, even their current energy level is more.

Table 2. Energy monitoring in aatral for the typical logistics sensor network

Aatral and smart logistics				
Nodetype	Node	GeoLocation	Current state	Energy level(%)
RFID	RF1-Store	Houston	Active	16
	RF2-Store	Houston	Sleep	12
	RF2-Warehouse	NewJersey	Active	45
	RF3-Warehouse	NewJersey	Sleep	87
	RF4-Container	LosAngles	Request	42
	RF4-Fleet1-Vehicle1	Philadelphia	Response	82
	RF5-Fleet1-Vehicle2	Phoenix	Active	60
	RF6-Fleet-Vehicle3	SanAntonio	Response	70
	RF7-Classification Center1	Austin	Active	80
	RF8-Classification Center2	Dallas	Sleep	76
	RF9-Distribution Center 1	SanJose	Response	92
	RF10-Distribution Center3	Detroit	Active	95
ImageSensor	IM1- Store	Houston	Active	43
	IM2-Store	Houston	Idle	50
	IM3-Warehouse	NewJersey	Active	25
	IM4-Classifiction Center1	Austin	Sleep	45
	IM5-Classification Center2	Dellas	Request	87
	IM6-Distribution Center 1	SanJose	Response	42

c. Energy Dissipation Model For Autonomous Logistics System. From the Table 3 and Fig. 14, it is obvious that the average number of events of the sensor is directly proportional to the energy dissipation rate. But, the number of events alone do not decide the dissipation rate, but it also depends on the distance of communication. As discussed, the trend rate of dissipation is calculated based on the historical data with a timestamp.

Figures 14 and 15 exhibit how it is possible to track individual sensor nodes energy dissipation as well as network energy dissipation in Aatral. The activity wise, event wise, unit wise tracking of energy consumption information is possible with Aatral.

Investigation of energy consumption of a sensor in the network is quite easy as all types of activities are measured and reported along with energy measures in this framework. The highlighted two rows in the Table 4 explains, that the dissipation rate or the trend of dissipation depends on the number of events as well as the distance of communication.

The baud rate, the protocol chosen, the mean coefficient of distribution of nodes, trend factor on energy dissipation, hardware of sensors and the drivers used all together help us to decide the expected lifetime of the sensor node. The automated workflow on every energy scan finds the energy level below the threshold and alert the consumed person or places the purchase order and raises the service request for battery replacement. So, the network coverage loss can be avoided proactively (Fig. 16).

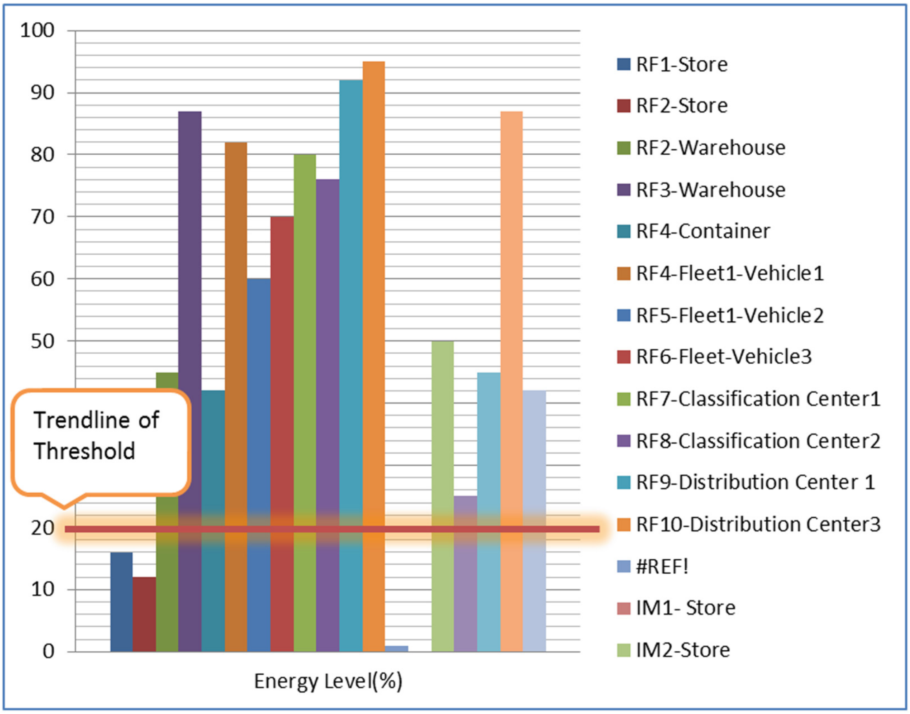


Fig. 13. Energy level monitoring of sensors with threshold trendline

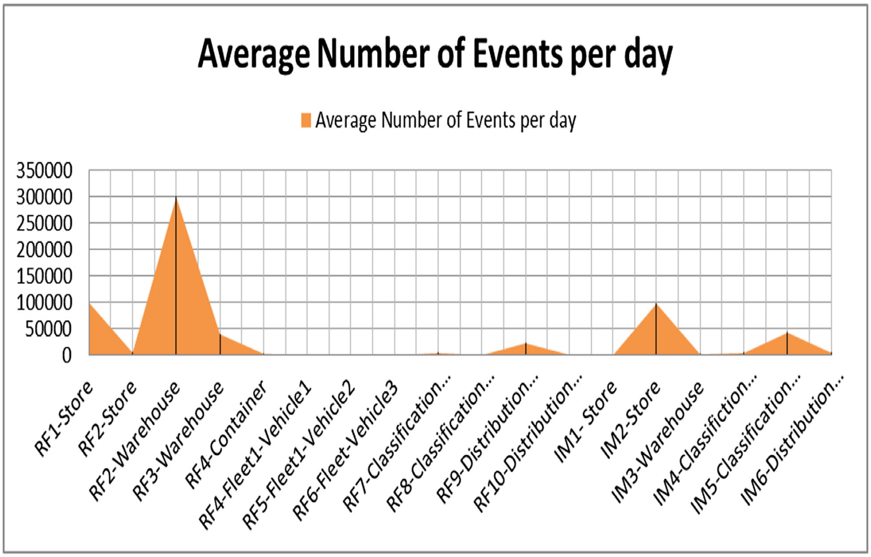


Fig. 14. Average number of event per day in the logistics sensor network

Table 3. Average number of events per day for sensors tracking

Smart logistics dashboard of Aatral				
Nodetype	Node	GeoLocation	Current state	Average number of events per day
RFID	RF1-Store	Houston	Active	100000
	RF2-Store	Houston	Sleep	5000
	RF2-Warehouse	NewJersey	Active	300000
	RF3-Warehouse	NewJersey	Sleep	40000
	RF4-Container	LosAngles	Request	3000
	RF4-Fleet1-Vehicle1	Philadelphia	Response	45
	RF5-Fleet1-Vehicle2	Phoenix	Active	23
	RF6-Fleet-Vehicle3	SanAntonio	Response	67
	RF7-Classification Center1	Austin	Active	4500
	RF8-Classification Center2	Dallas	Sleep	567
	RF9-Distribution Center 1	SanJose	Response	23000
	RF10-Distribution Center3	Detroit	Active	1200
Image Sensor	IM1- Store	Houston	Active	23
	IM2-Store	Houston	Idle	98000
	IM3-Warehouse	NewJersey	Active	2000
	IM4-Classifiction Center1	Austin	Sleep	4500
	IM5-Classification Center2	Dellas	Request	43000
	IM6-Distribution Center 1	SanJose	Response	5000

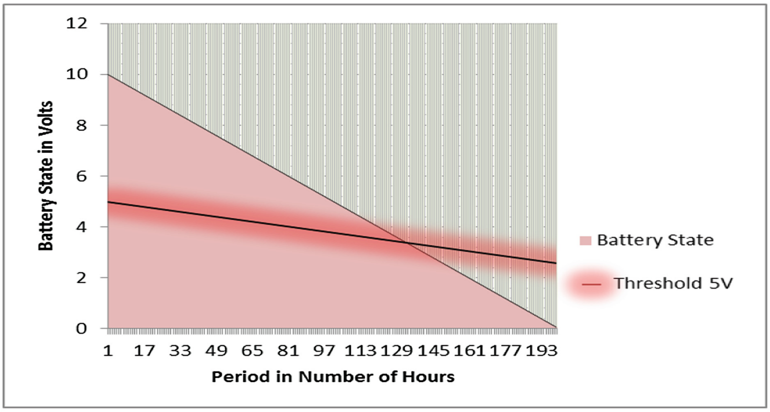


Fig. 15. Battery state of a sensor over the period of time.

Table 4. Life time prediction of sensors in smart logistics

Aatral and smart logistics						
Node type	Node	Energy level (%)	Average number of events per day	Average distance of communication per day in Km	Dissipation rate in %	Life time prediction in days
RFID	RF1-Store	16	100000	0.25	0.25	0.64
	RF2-Store	12	5000	0.65	0.0325	3.692307692
	RF2-Warehouse	45	300000	0.52	1.56	0.288461538
	RF3-Warehouse	87	40000	0.75	0.3	2.9
	RF4-Container	42	3000	12	0.36	1.166666667
	RF4-Fleet1-Vehicle1	82	45	340	0.153	5.359477124
	RF5-Fleet1-Vehicle2	60	23	1200	0.276	2.173913043
	RF6-Fleet-Vehicle3	70	67	200	0.134	5.223880597
	RF7-Classification Center1	80	4500	400	18	0.044444444
	RF8-Classification Center2	76	567	600	3.402	0.223398001
	RF9-Distribution Center 1	92	23000	23	5.29	0.173913043
	RF10-Distribution Center3	95	1200	120	1.44	0.659722222
Image sensor	IM1- Store	43	23	0.25	0.0000575	7478.26087
	IM2-Store	50	98000	0.65	0.637	0.784929356
	IM3-Warehouse	25	2000	0.72	0.0144	17.36111111
	IM4-Classifiction Center1	45	4500	0.5	0.0225	20
	IM5-Classification Center2	87	43000	45	19.35	0.04496124
	IM6-Distribution Center 1	42	5000	23	1.15	0.365217391

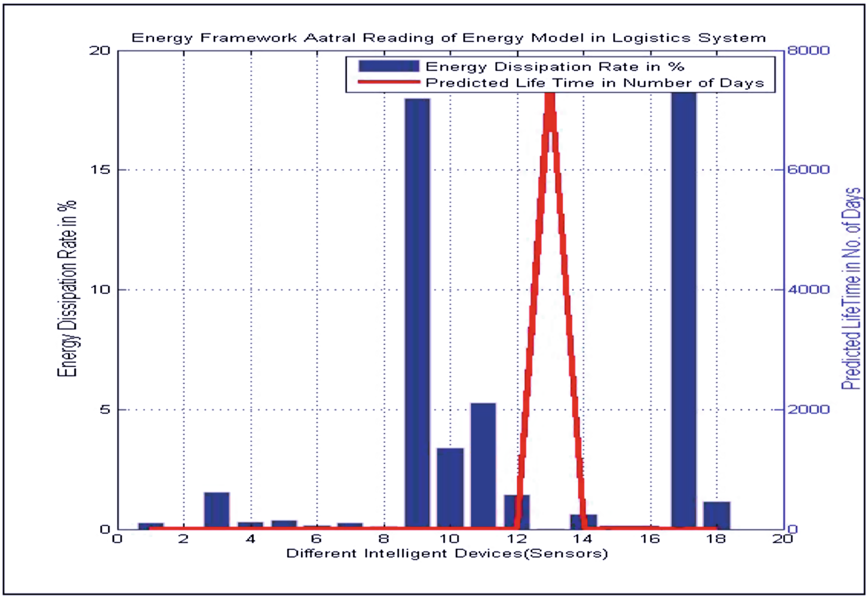


Fig. 16. Energy level of sensors (1..18) and the average of number of events

d. Comparison of Life Time Prediction Models. The probability based models and random variable based models are considered here for the comparison of real world accurate energy model with the trend factor [19, 20]. In all the theoretical models, the computational overhead of distance calculation and the angle calculation for estimating the energy for propagation have the impact on the performance (Table 5).

Table 5. Comparison of different life time prediction approaches

Aatral life time prediction model out performs in smart logistics			
Factors	Accuracy in probability based model of life time prediction in %	Accuracy in random model of life time prediction %	Accuracy in life time prediction with energy model and trend factor %
Protocols			
Zigbee	66	76	96
IEEE 802.15.4	72	87	97
MIWI (short range)	82	89	94
Baud rate			
120	76	84.5	98
1200	65	73	95
12000	62	69	92
64	84	74	96
128	82	74	95.5
256	85	74	94.3
Performance			
Throughput	72	69	90
Latency	32	43	12
Collision and retransmission rate	21	18	7
Channel utilization	64	72	94

Ultimately, the energy scan, energy auditing and energy reporting should not affect the fairness of the Quality Of Service (QOS) parameters. It has been validated with the throughput, channel utilization, latency delay for different baud rates and protocols with the Aatral framework (Fig. 17).

11 Summary of Findings and Real World Implications

The findings can be summarized as follows:

a. In the autonomous cooperative smart logistics the whole process is getting automated with the auto flow of sensor data. Loss of energy in the sensor may disconnect the sensor and leads to not only the loss of coverage, but also stops the data inflow which will affect the workflow configured. The 'Aatral' energy management framework proposes the potential energy scan on a frequent interval and reports the

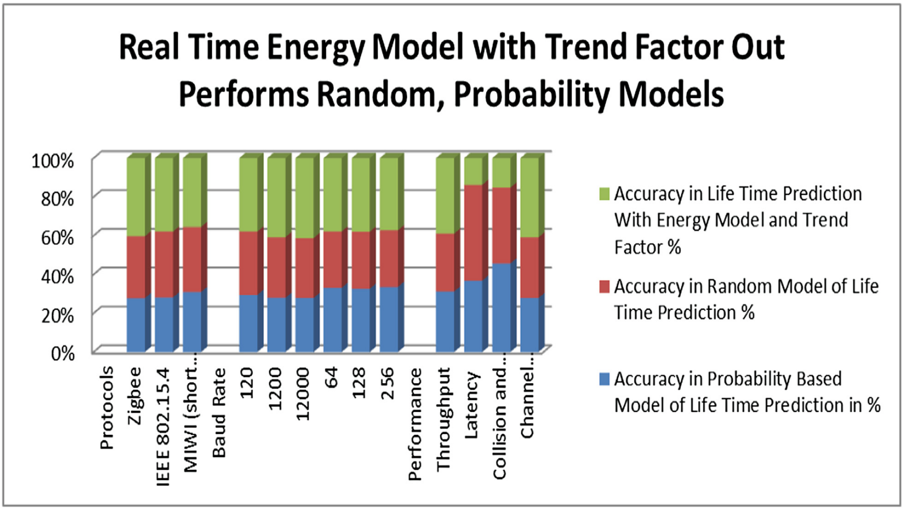


Fig. 17. Comparison of life time prediction mechanisms

energy map of the sensors in the logistics network. The workflow is configured to propagate the alert or initiate the purchase order or raise the service request to replace batteries. It is easy to manage the logistics system with the right escalation mechanism configuration.

b. Considering the baud rate, protocol selection, hardware of the platform, average number of events, average distance of communication and the historical data over timestamp, a trend factor is calculated and based on that, the energy dissipation rate is derived.
c. The energy dissipation rate and the current energy level help to predict the lifetime of the sensors. By this predictive mechanism, the autonomous logistics system can be operational with zero downtime with the full coverage of a network.

12 Conclusion and Future Directions

There is a paradigm shift from the centralized logistics to autonomous cooperative logistics with the intelligent devices and distributed systems. Here the energy of the intelligent devices is limited. So, there must be a predictive energy mechanism for the energy measurements of the intelligent devices so that there won't be the coverage loss which will be very crucial in the smart logistics environment. In future, the service users can configure the service level agreements (SLA), not only the delivery preferences, but also the energy preferences and possible routes logics based on the energy and the energy escalation matrices, energy indications and alerts on battery replacement or alternate energy sources. So, the knowledge representation must include the "energy data" of the IOT in the autonomous logistics to have zero down time and energy efficient autonomous operations. Trend factor is also based on the historical data and other factors like baud rate. In some cases, when the sensor profile changes or the service change configuration in the usage,they won't have the immediate affect the trend factor that could be attributed as the slippage of the accuracy in prediction even though by a slight margin only. Later in the IOT platforms, using the big data techniques, the whole operation of the predictive mechanism can be scaled.

Acknowledgement. We would like to thank anonymous reviewers whose careful reading and constructive criticism of earlier draft helped to improve the clarity and content of this paper.

References

1. http://www.usa.gov/directory/federal/department-of-transportation.shtml
2. Le, C.V., Pang, C.K.: An energy data-driven decision support system for high-performance manufacturing industries. Int. J. Autom. Logistics **1**(1), 61–79 (2013)
3. Sivamani, S., Kwak, K., Cho, Y.: A study on intelligent user-centric logistics service model using ontology. J. Appl. Math., **2014**(162838) (2014)
4. Baars, H., Kemper, H.-G., Lasi, H., Siegel, M.: Combining RFID technology and business intelligence for supply chain optimization – scenarios for retail logistics. In: Proceedings of the 41st Hawaii International Conference on System Sciences (2008)
5. De Laeta, V., van Loonb, G., Van der Perreb, A.: Integrated remote sensing investigations of ancient quarries and road systems in the Greater Dayr al-Barshā Region Middle Egypt: a study of logistics. J. Archaeol. Sci. **55**, 286–300 (2015)
6. Abad, E., Zampolli, S., Marco, S.: Flexible tag microlab development: gas sensors integration in RFID flexible tags for food logistics, 6 EADS Deutschland GmbH, Corporate Research Centre, München, Germany
7. Monostori, L., Valckenaers, P., Dolgui, A.: Cooperative control in production and logistics, The International Federation of Automatic Control Cape Town, South Africa, 24–29 August 2014
8. Vieira, C.V.A., Cardoso, A.J.M.: The role of information logistics and data warehousing in educational facilities asset management. Int. J. Syst. Assur. Eng. Manag. **1**(3), 229–238 (2010)
9. Mankiw, N.G.: Smart taxes: an open invitation to join the Pigou club. Eastern Econ. J. **35**, 14–23 (2009)
10. Hu, Z.-H., Yang, B., Huang, Y.-F.: Visualization framework for container supply chain by information acquisition and presentation technologies, J. Softw., **5**(11), November 2010
11. Thangaraj., M, Anuradha, S.: Setting up an energy measurable application bed of wireless sensor network for the improved energy economics. In: 6th International Conference on Advanced Computing, MIT, Chennai (2014)
12. Sharma, T., Kelkar, D.: A Tour Towards Knowledge Representation Techniques. Int. J. Comput. Technol. Electron. Eng. (IJCTEE), **2**(2). ISSN 2249–6343
13. Luo, J., Jiang, L.G., He, C.: An analytical model for SMAC protocol in multi-hop wireless sensor networks. Sci. Chin. Inf. Sci. **53**(11), 2323–2331 (2010)
14. Berger, J.O., de Oliviera, V., Sanso, B.: Objective bayesian analysis of spatially correlated data. J. Am. Stat. Assoc. **96**, 1361–1374 (2001)
15. Mitra, S.K., Naskar, M.K.: Comparative study of radio models for data gathering in wireless sensor network. Int. J. Comput. Appl. **27**(4), 49–57 (2011)
16. Int. J. Comput. Appl., **27**(4), August 2011. ISSN 0975 – 8887

17. Wang, X., Ma, J.J., Wang, S., Bi, D.W.: Prediction based dynamic energy management in wireless sensor networks sensor networks. Sens. J. **7**(3), 251–266 (2007)
18. Wang, X., Ma, J.J., Wang, S., Bi, D.W.: Cluster-based dynamic energy management for collaborative target tracking in wireless sensor networks. Sens. J. **7**(7), 1193–1215 (2007)
19. Medeiros, H., Park, J., Kak, A.C.: Distributed object tracking using a cluster-based kalman filter in wireless camera networks. IEEE J. Sel. Top. Sig. Process. **2**(4), 448–463 (2008)
20. Prajapat, M., Barwar, N.C.: Performance analysis of energy dissipation in WSNs using multi-chain PEGASIS. Int. J. Comput. Sci. Inf. Technol. **5**(6), 8033–8036 (2014)
21. Jiang, H., Wuhan,Jin, S.: Prediction or not? an energy-efficient framework for clustering-based data collection in wireless sensor networks. IEEE Transactions Parallel and Distributed Systems, **22**(6)

A Study on the Effects of LED Light's Pulse Width Modulation on Work Concentration

Hyung-sun Kim, Gen-Soo Song, Hyo-Beom Ahn, Chang-Yong Lee, and Jin Ho Kim(✉)

Advanced Informatics School (AIS), Universiti Teknologi Malaysia (UTM), 54100 Kuala Lumpur, Malaysia
ddaker98@naver.com, ds-9807@daum.net, {hbahn,clee,kjh}@kongju.ac.kr

Abstract. This study examines the effects of changes in LED lighting on a subject's work environment by varying the characteristics of LED light provided to an experimental work environment while subjects complete an error correcting activity. The study configured a variety of lighting environments using an LED lighting control system, through which the control of pulse width modulation (PWM) and illuminance can be conducted. For the indoor environment of the test bed, the comfort index in winter recommended by ASHRAE Standard is applied. The nine types of lighting environment are configured, according to PWM and illuminance. Concerning the configured lighting environment, the PWM ratio is set as R:G:B = 1:1:1, R:G:B = 4:1:5, R:G:B = 8:7:7, and illuminance is set as 400 lx, 700 lx, 1000 lx. This study also analyzes the change of subject concentration levels through real time measurement of brainwaves, while the error search and correction is carried out in the suggested lighting environment. This was analyzed through relative SMR analysis. The experiment did not find significant differences in concentration levels resulting from the different lighting conditions.

Keywords: Pulse width modulation · Illuminance · Electroencephalogram · Relative SMR

1 Introduction

Many national policy projects have recently been carried out in order to prevent the depletion of energy and to promote environmental conservation worldwide. Among others, many studies on LED lighting have been conducted for energy use reduction and to address environmental problems. As LED lighting developed, some measures of the impact of various aspects of the lighting are required.

Since people encounter lighting mostly in leading indoor life, studies on the standards and evaluation of comfort levels resulting from changes of lighting are often performed. Those studies suggested that various lighting factors need to be reflected,

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 550–560, 2015.
DOI: 10.1007/978-3-319-21009-4_41

according to the analysis on satisfaction felt by workers on lighting in working spaces and the characteristics of workers [7, 12]. There are some studies on the development of objective evaluation tools for visual work and fatigue, and objective evaluation result analyses, according to the change of lighting environment [1, 9, 11]. This implies that studies on lighting environment change for work efficiency are required.

Some studies to analyze human's bio signals to quantitatively evaluate comfort on lighting environmental change have been conducted, and many studies on the brainwave analysis among human bio signals have been carried out. Some performed studies on the 12 ~ 15 Hz SMR wave revelation indicating attention concentration, which is an intermediate process between alpha and low beta waves [6, 8]. Some studies analyzed attention concentration with 12 ~ 15 Hz SMR wave activity indicating 4 ~ 8 Hz theta wave's activity and attention concentration, which also shows the state of drowsiness [8], or analyzed middle beta wave activity ratio indicating the activity and disillusion of theta and SMR waves [3, 4]. [2, 10] conducted studies on bio feedback interfaces through which users' concentration and emotional stage can be recognized through brainwave and electrocardiogram analysis. If a subject's state can be identified by measuring and analyzing bio signals related with attention concentration for work processing in real time, the result can be used to configure a comfortable work environment. However, existing studies are mostly for illuminance, color temperature and colors of LED light and the studies on the characteristics of lighting through pulse width modulation (PWM) are insufficient.

This study aims to analyze the effects of change in the lighting environment on work efficiency through PWM, which is one of the characteristics of LED light. This study built a test bed similar to an indoor work space and developed an LED light system that can control PWM. Nine types of various lighting environments were configured using the LED lighting system. This study carried out error search and correction in each lighting environment to analyze the effects on work efficiency and measured brainwaves in real time. This study actually identifies the effects on user's indoor work by analyzing the measured brainwaves during an error search and correction task.

2 Methodology

This study built a test bed similar to an indoor work environment, with LED stand lighting that could be remotely controlled for pulse and luminance. To determine changes in user's work efficiency related to changes in the lighting environment the test subjects were asked to perform error search and correction tasks and the study analyzed changes in concentration levels by measuring brainwaves during the activity.

2.1 Experimental Environment and Subject

The experimental space was divided into an area for the experiment and a preparation area. Figure 1 shows the entire test bed structure. For the experimental space, blackout curtains were set up to interrupt light introduced from the outside. Temperature was

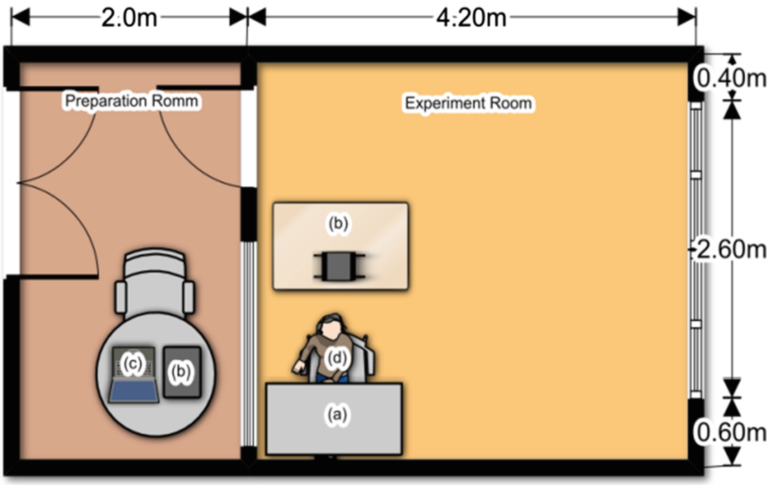


Fig. 1. Test bed((a) Working table (b and c) Bio-signal measurement device (d) Subject)

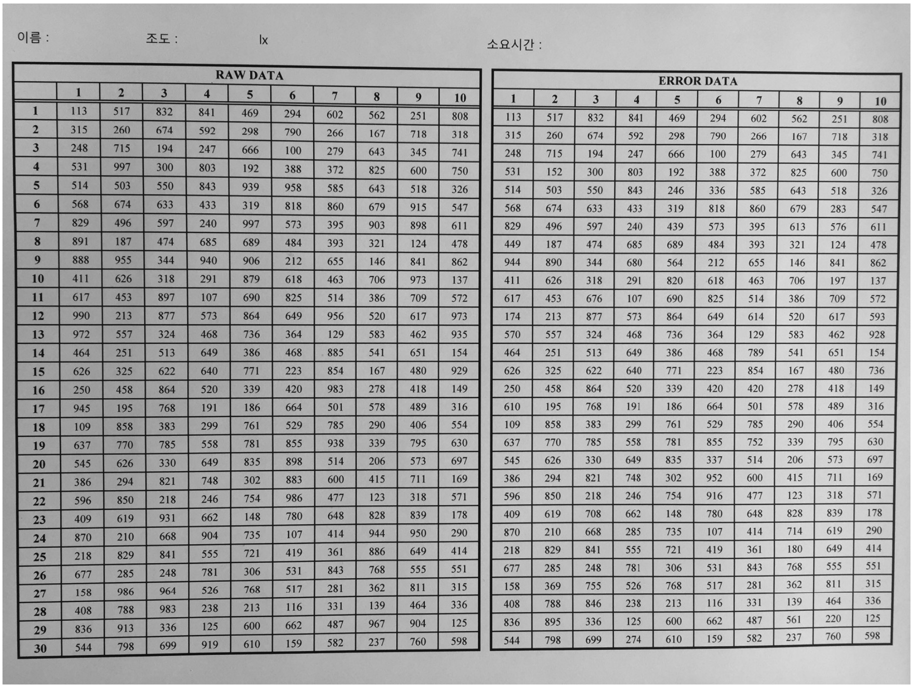
controlled through cooling and heating devices, while humidity was maintained by installing a dehumidifier.

A PMV comfort index of 0 (as recommended by SHHRAE standard) was maintained for the indoor environment, and the experiment was carried out by maintaining room temperature at 20−24°C, 50−60 % humidity and 1 of clothing. Temperature and humidity were measured at the height of 1.2 m in the central indoor part through a KEM's AM-101 device in real time in order to maintain the indoor environment. For the lighting environment, changes to illuminance was made by differentiating the LED lighting's pulse ratio. The applied PWM ratios were R:G:B = 1:1:1, R:G:B = 4:1:5, R:G:B = 8:7:7, and three types of different illuminance, 400 lx, 700 lx, 1000 lx were set. Each lighting environment was remote controlled through the LED lighting control system made in this study.

The subjects participating in this study were ten physically and mentally healthy males (age: 23.5 ± 0.81) and ten females (age: 21 ± 0.89). They did not have any history of neuropsychiatric disorders including brain trauma, epilepsy and alcoholism, and they received pre-education on the experiment. One day before the experiment, the subjects were instructed to avoid foods such as alcoholic drinks, and caffeinated beverages that affect cognitive function.

2.2 Evaluation Tools

To evaluate the indoor work environment efficiency suggested in this study, this study modified tools to be suitable for the level of the students by using the "Data Search and Correction Task" evaluation tool developed by the Korea Standard Research Institute. Three error search and correction evaluation tools were used, and produced on A4 paper as shown in Fig. 2 to Fig. 4.



이름 : 조도 : lx 소요시간 :

RAW DATA	1	2	3	4	5	6	7	8	9	10
1	113	517	832	841	469	294	602	562	251	808
2	315	260	674	592	298	790	266	167	718	318
3	248	715	194	247	666	100	279	643	345	741
4	531	997	300	803	192	388	372	825	600	750
5	514	503	550	843	939	958	585	643	518	326
6	568	674	633	433	319	818	860	679	915	547
7	829	496	597	240	997	573	395	903	898	611
8	891	187	474	685	689	484	393	321	124	478
9	888	955	344	940	906	212	655	146	841	862
10	411	626	318	291	879	618	463	706	973	137
11	617	453	897	107	690	825	514	386	709	572
12	990	213	877	573	864	649	956	520	617	973
13	972	557	324	468	736	364	129	583	462	935
14	464	251	513	649	386	468	885	541	651	154
15	626	325	622	640	771	223	854	167	480	929
16	250	458	864	520	339	420	983	278	418	149
17	945	195	768	191	186	664	501	578	489	316
18	109	858	383	299	761	529	785	290	406	554
19	637	770	785	558	781	855	938	339	795	630
20	545	626	330	649	835	898	514	206	573	697
21	386	294	821	748	302	883	600	415	711	169
22	596	850	218	246	754	986	477	123	318	571
23	409	619	931	662	148	780	648	828	839	178
24	870	210	668	904	735	107	414	944	950	290
25	218	829	841	555	721	419	361	886	649	414
26	677	285	248	781	306	531	843	768	555	551
27	158	986	964	526	768	517	281	362	811	315
28	408	788	983	238	213	116	331	139	464	336
29	836	913	336	125	600	662	487	967	904	125
30	544	798	699	919	610	159	582	237	760	598

ERROR DATA									
1	2	3	4	5	6	7	8	9	10
113	517	832	841	469	294	602	562	251	808
315	260	674	592	298	790	266	167	718	318
248	715	194	247	666	100	279	643	345	741
531	152	300	803	192	388	372	825	600	750
514	503	550	843	246	336	585	643	518	326
568	674	633	433	319	818	860	679	283	547
829	496	597	240	439	573	395	613	576	611
449	187	474	685	689	484	393	321	124	478
944	890	344	680	564	212	655	146	841	862
411	626	318	291	820	618	463	706	197	137
617	453	676	107	690	825	514	386	709	572
174	213	877	573	864	649	614	520	617	593
570	557	324	468	736	364	129	583	462	928
464	251	513	649	386	468	789	541	651	154
626	325	622	640	771	223	854	167	480	736
250	458	864	520	339	420	420	278	418	149
610	195	768	191	186	664	501	578	489	316
109	858	383	299	761	529	785	290	406	554
637	770	785	558	781	855	752	339	795	630
545	626	330	649	835	337	514	206	573	697
386	294	821	748	302	952	600	415	711	169
596	850	218	246	754	916	477	123	318	571
409	619	708	662	148	780	648	828	839	178
870	210	668	285	735	107	414	714	619	290
218	829	841	555	721	419	361	180	649	414
677	285	248	781	306	531	843	768	555	551
158	369	755	526	768	517	281	362	811	315
408	788	846	238	213	116	331	139	464	336
836	895	336	125	600	662	487	561	220	125
544	798	699	274	610	159	582	237	760	598

Fig. 2. Number error search and correction

Number error search and correction: The number error search and correction evaluation sheet was a table composed of 10 columns and 30 rows by classifying Table A and Table B on the A4 paper as shown in Fig. 2. The font is Times New Roman, and font size is 9 point. Concerning evaluation details, three-digit numbers from 100 to 999 are randomly arranged, and the number of wrong numbers included is 40, and they are also randomly arranged.

Letter error search and correction: The number error search and correction evaluation sheet was divided into two tables (left and right) on an A4 paper as shown in Fig. 3, and is composed of 20 columns and 25 rows, respectively. Font is Gothic and font size is 9 point. Concerning evaluation details, one-letter Korean characters are randomly arranged, and the number of wrong characters included is 40 and they are also randomly arranged.

Arithmetic operation error search and correction: The arithmetic operation number error search and correction evaluation sheet was the table composed of 4 columns and 40 rows on A4 paper as shown in Fig. 4. Font is Times New Roman and font size is 9 point. Concerning evaluation details, single-digit number arithmetic operations are randomly arranged, and the number of arithmetic operation errors included is 40 and they are also randomly arranged.

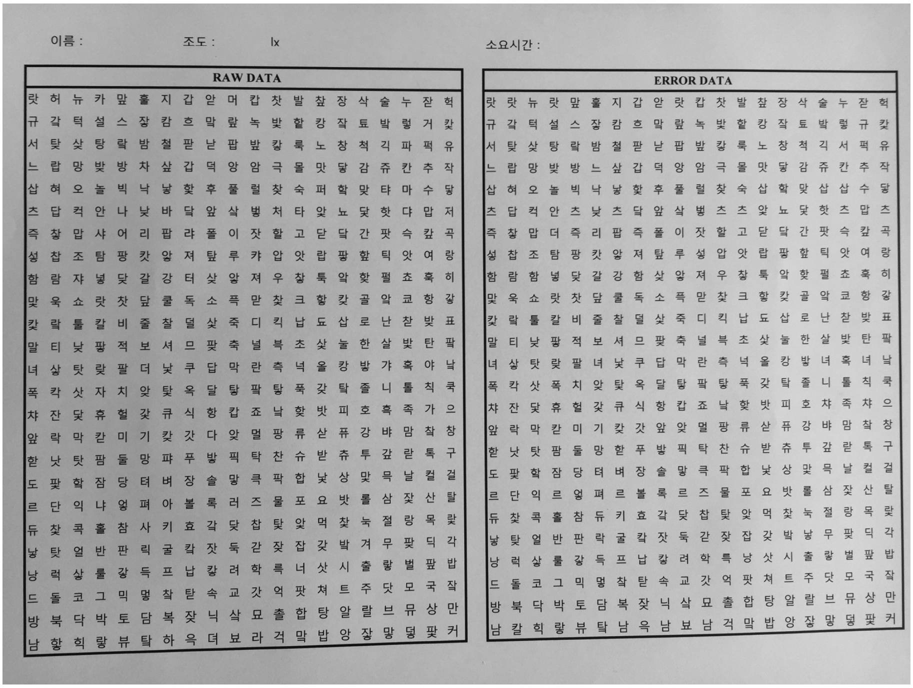


Fig. 3. Letter error search and correction

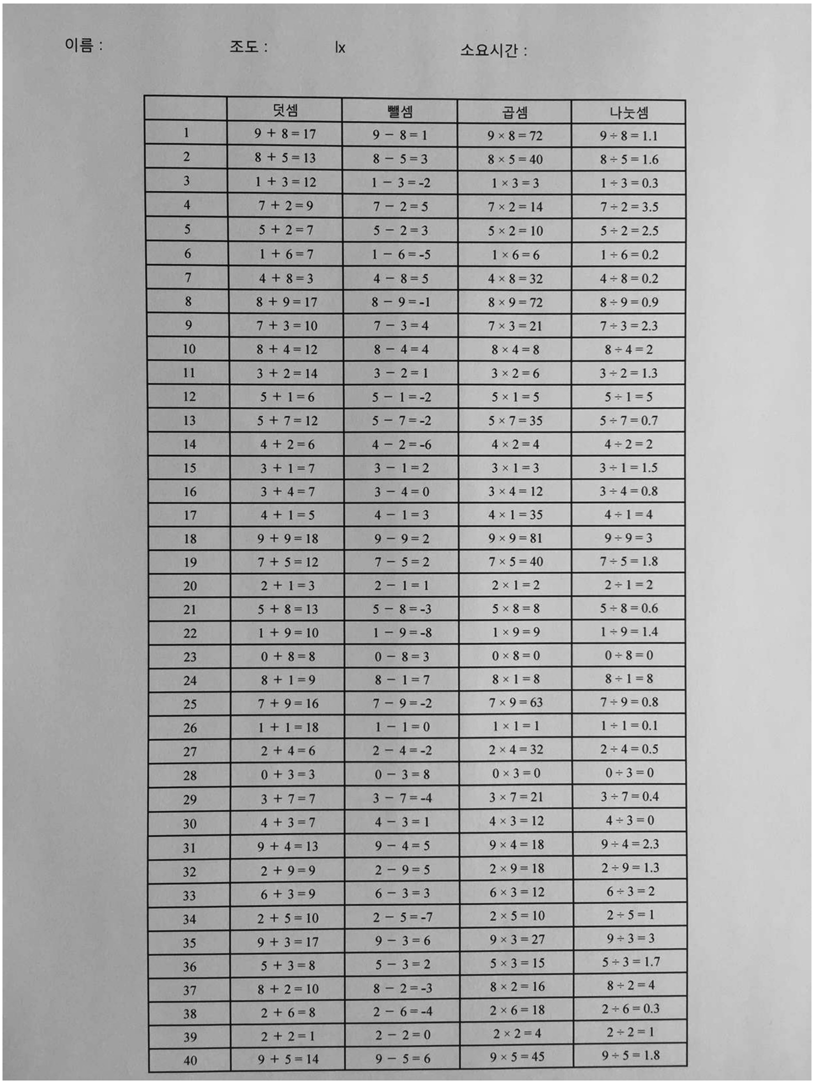
Hypothesis 3: Access control has a positive relationship with privacy preferences in healthcare environment.

2.3 Bio-Signals Measuring System

To measure the subject's bio-signals, namely, brainwaves, this study used a PolyG-I system and TeleScan 3.09 program of LAXTHA as shown in Fig. 5.

According to the international 10/20 electrode system (Jasper, 1958) for brainwave measurement, electrodes were located on F_{p1} and F_{p2}, which are prefrontals, and monopolar derivation, based on the underneath of the right ear, was used. In the experiment, cautions on maximum motion area scope and movement were provided to the subjects to minimize artifacts, when the subjects conducted the error search and correction. Brainwave data collection was remotely conducted in the experiment preparatory space. For the brainwave data measured through this experiment, this study analyzed the relative ratio values, in which the vibration of the area concerned out of the entire area appeared, through the relative power method.

$$\text{Relative power} = \frac{x_{power}}{\theta_{power} + \alpha_{power} + \beta_{power} + \gamma_{power}} \quad (1)$$



이름 : 조도 : lx 소요시간 :

	덧셈	뺄셈	곱셈	나눗셈
1	9 + 8 = 17	9 − 8 = 1	9 × 8 = 72	9 ÷ 8 = 1.1
2	8 + 5 = 13	8 − 5 = 3	8 × 5 = 40	8 ÷ 5 = 1.6
3	1 + 3 = 12	1 − 3 = -2	1 × 3 = 3	1 ÷ 3 = 0.3
4	7 + 2 = 9	7 − 2 = 5	7 × 2 = 14	7 ÷ 2 = 3.5
5	5 + 2 = 7	5 − 2 = 3	5 × 2 = 10	5 ÷ 2 = 2.5
6	1 + 6 = 7	1 − 6 = -5	1 × 6 = 6	1 ÷ 6 = 0.2
7	4 + 8 = 3	4 − 8 = 5	4 × 8 = 32	4 ÷ 8 = 0.2
8	8 + 9 = 17	8 − 9 = -1	8 × 9 = 72	8 ÷ 9 = 0.9
9	7 + 3 = 10	7 − 3 = 4	7 × 3 = 21	7 ÷ 3 = 2.3
10	8 + 4 = 12	8 − 4 = 4	8 × 4 = 8	8 ÷ 4 = 2
11	3 + 2 = 14	3 − 2 = 1	3 × 2 = 6	3 ÷ 2 = 1.3
12	5 + 1 = 6	5 − 1 = -2	5 × 1 = 5	5 ÷ 1 = 5
13	5 + 7 = 12	5 − 7 = -2	5 × 7 = 35	5 ÷ 7 = 0.7
14	4 + 2 = 6	4 − 2 = -6	4 × 2 = 4	4 ÷ 2 = 2
15	3 + 1 = 7	3 − 1 = 2	3 × 1 = 3	3 ÷ 1 = 1.5
16	3 + 4 = 7	3 − 4 = 0	3 × 4 = 12	3 ÷ 4 = 0.8
17	4 + 1 = 5	4 − 1 = 3	4 × 1 = 35	4 ÷ 1 = 4
18	9 + 9 = 18	9 − 9 = 2	9 × 9 = 81	9 ÷ 9 = 3
19	7 + 5 = 12	7 − 5 = 2	7 × 5 = 40	7 ÷ 5 = 1.8
20	2 + 1 = 3	2 − 1 = 1	2 × 1 = 2	2 ÷ 1 = 2
21	5 + 8 = 13	5 − 8 = -3	5 × 8 = 8	5 ÷ 8 = 0.6
22	1 + 9 = 10	1 − 9 = -8	1 × 9 = 9	1 ÷ 9 = 1.4
23	0 + 8 = 8	0 − 8 = 3	0 × 8 = 0	0 ÷ 8 = 0
24	8 + 1 = 9	8 − 1 = 7	8 × 1 = 8	8 ÷ 1 = 8
25	7 + 9 = 16	7 − 9 = -2	7 × 9 = 63	7 ÷ 9 = 0.8
26	1 + 1 = 18	1 − 1 = 0	1 × 1 = 1	1 ÷ 1 = 0.1
27	2 + 4 = 6	2 − 4 = -2	2 × 4 = 32	2 ÷ 4 = 0.5
28	0 + 3 = 3	0 − 3 = 8	0 × 3 = 0	0 ÷ 3 = 0
29	3 + 7 = 7	3 − 7 = -4	3 × 7 = 21	3 ÷ 7 = 0.4
30	4 + 3 = 7	4 − 3 = 1	4 × 3 = 12	4 ÷ 3 = 0
31	9 + 4 = 13	9 − 4 = 5	9 × 4 = 18	9 ÷ 4 = 2.3
32	2 + 9 = 9	2 − 9 = 5	2 × 9 = 18	2 ÷ 9 = 1.3
33	6 + 3 = 9	6 − 3 = 3	6 × 3 = 12	6 ÷ 3 = 2
34	2 + 5 = 10	2 − 5 = -7	2 × 5 = 10	2 ÷ 5 = 1
35	9 + 3 = 17	9 − 3 = 6	9 × 3 = 27	9 ÷ 3 = 3
36	5 + 3 = 8	5 − 3 = 2	5 × 3 = 15	5 ÷ 3 = 1.7
37	8 + 2 = 10	8 − 2 = -3	8 × 2 = 16	8 ÷ 2 = 4
38	2 + 6 = 8	2 − 6 = -4	2 × 6 = 18	2 ÷ 6 = 0.3
39	2 + 2 = 1	2 − 2 = 0	2 × 2 = 4	2 ÷ 2 = 1
40	9 + 5 = 14	9 − 5 = 6	9 × 5 = 45	9 ÷ 5 = 1.8

Fig. 4. Arithmetic operation error search and correction

2.4 Experiment Method

This experiment was conducted using two-way ANOVA containing repetition. In the experiment, the subjects carried out error search and correction tasks under three PWM and three LED illumination conditions, and the corresponding brainwaves were measured. In doing so, experimental conditions were selected randomly. Here, PWM conditions were PWM_1 (R:G:B = 1:1:1), PWM_2 (R:G:B = 4:1:5), and PWM_3 (R:G:B = 8:7:7) and LED illuminance conditions were 400 lx, 700 lx, and 1000 lx. After the

Fig. 5. PolyG-I system

experiment was completed for all condition combinations for one subject, the experiment was conducted for another subject.

The specific experiment procedure for one subject is as follows: Adaptation to indoor environment and electrode location (10 min) → Dark adaptation (2 min) → Light adaptation (2 min) → Error search and correction and bio-signals measurement (20 min) → Dark adaptation (2 min) → Light adaptation (2 min) → Error search and correction and bio-signals measurement (20 min) → Dark adaptation (2 min) → Light adaptation (2 min) → Error search and correction and bio-signals measurement (20 min) → Electrode removal and tidying up (5 m).

3 Result

For the experiment, subjects performed error search and correction activities under each type of lighting condition while their brainwaves were being measured in real time. The brainwaves were collected at two channels, F_{p1} and F_{p2}, and this study analyzed the concentration change of subjects through relative SMR analysis. To identify the significant differences of brainwave data, this study used Minitab 16 for Windows.

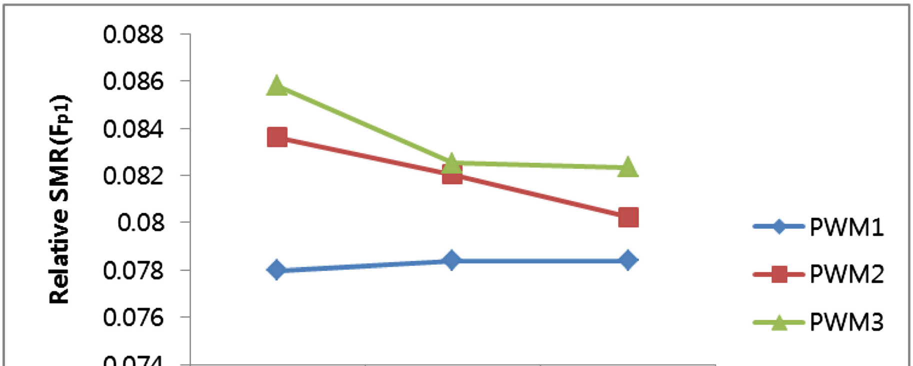
3.1 Analysis of Concentration on F_{p1} Channel

Looking at Table 1, the ANOVA result on the concentration of brainwave (F_{p1}), there was no significant difference between PWM values (p value = 0.236), and there was no significant difference between LEDs (illuminance) (p value = 0.791). There was no significant difference between PWM and LED in terms of reciprocal action (p value = 0.982).

Figure 6 shows the graph of average concentration analyzed through relative SMR on channel F_{p1}. For all LEDs (illuminance), there was change of concentration in the order of $PWM_3 > PWM_2 > PWM_1$, and the PWM_3 condition showed the highest concentration in all LEDs (illuminance). The highest concentration was revealed in PWM_3, and 400 lx out of all types of LEDs (illuminance) environment.

Table 1. Result of ANOVA on relative SMR of EEG (F_{p1})

Source	SS	DF	MS	F	P
PWM	0.0007097	2	0.0003548	1.46	0.236
LED	0.0001144	2	0.0000572	0.24	0.791
PWM*LED	0.0000984	4	0.0000246	0.10	0.982
Error	0.0328031	135	0.0002430		
Total	0.0337255	143			



	400lx	700lx	1000lx
PWM1	0.077978063	0.078371614	0.078395152
PWM2	0.08359864	0.082031005	0.080237058
PWM3	0.085785412	0.082513198	0.082342637

Fig. 6. Result of average on relative SMR of EEG (F_{p1})

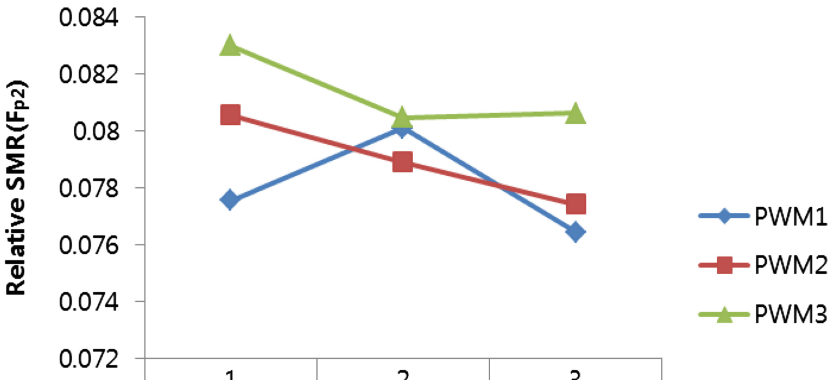
Table 2. Result of ANOVA on relative SMR of EEG (F_{p2})

Source	SS	DF	MS	F	P
PWM	0.0002836	2	0.0001418	0.43	0.654
LED	0.0001266	2	0.0000633	0.19	0.827
PWM*LED	0.0001283	4	0.0000321	0.10	0.983
Error	0.0448838	135	0.0003325		
Total	0.0454223	143			

3.2 Analysis of Concentration on F_{p2} Channel

Looking at Table 2, the ANOVA result on concentration of brainwave (F_{p2}), there was no significant difference between PWM values (p value = 0.654), and there was no significant difference between LEDs (illuminance) (p value = 0.827). There was no significant difference between PWM and LED in terms of reciprocal action (p value = 0.983).

Figure 7 shows the graph of average concentration analyzed through relative SMR on channel F_{p2}. For LEDs (illuminance), 400 lx and 1000 lx, there was change of concentration in the order of $PWM_3 > PWM_2 > PWM_1$. In 700 lx, there was change of concentration in the order of $PWM_3 > PWM_1 > PWM_2$. For all LEDs (illuminance), PWM_3 condition showed the highest concentration. The highest concentration was revealed in PWM_3, and 400 lx out of all types of LED (illuminance) environment.



	1	2	3
PWM1	0.077542652	0.080085101	0.076421902
PWM2	0.080543205	0.078888899	0.077409589
PWM3	0.082972156	0.080456034	0.080615402

Fig. 7. Result of average on relative SMR of EEG (F_{p2})

4 Conclusion

This study aimed to identify the effects of pulse width modulation and change in illuminance, which are the characteristics of lighting in a comfortable indoor environment defined in ASHRAE standards, on a user's work efficiency. To this end, an experiment was undertaken by configuring various types of lighting environment with zero PMV value in winter in terms of indoor environment. Three types of pulse width modulation (PWM) of LED light were presented for the lighting environment, and 400 lx, 700 lx, and 1000 lx were presented in terms of the brightness of LED (illuminance): nine types of environment were offered in total. For each lighting environment, error search and correction tasks were conducted, and work efficiency was identified. This study analyzed the users' concentration change through real time brainwave measurement.

Analysis of brainwaves did not show significant differences in concentration levels resulting from changes in PWM or LED (illuminance) in the two channels (F_{p1}, F_{p2}). However, looking at the changes recorded, the highest concentration levels were seen on change of each LED (illuminance). The highest concentration was shown in 400 lx in terms of LED (illuminance). Although, no significant difference was shown in all types of LED (illuminance) environment, the experiment was carried out in limited lighting environment. For this reason, further study is considered to be required supplementing the lighting environment.

Acknowledgments. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2009-0093825). Also, it was supported by the Power Generation & Electricity Delivery Core Technology Program of the Korea Institute of Energy Technology Evaluation and Planning (KETEP) granted financial resource from the Ministry of Trade, Industry & Energy, Republic of Korea (No. 20131020400850).

References

1. Yang, H.K., Ko, H.W., Kim, M.H., Lim, S.K., Yun, Y.H.: Evaluation of fatigue by analysis of relation between subjective rating score and working performance with color temperature. J. Korean Soc. Emot. Sensibility **4**(2), 63–68 (2001)
2. Kim, J.S., Kim, H.K., Jeong, H., Kim, K.H., Im, S.H., Son, W.H.: Human-computer interface based on bio-signal. Electron. Telecommun. Trends **20**(4), 67–81 (2005)
3. Jang, S.W., Lee, H.C.: The effect of musical genre on mental focusing. In: Proceeding of the Korean Psychological Association, 262–263 (2008)
4. Lee, C.H., Kwon, J.W., Kim, K.D., Hong, J.E., Shin, D.S., Lee, D.H.: A study on EEG based concentration transmission and brain computer interface application. Korean J. Inst. Electron. Inf. Eng. **46**(2), 41–46 (2009)
5. Shim, J.Y.: Influences of applying brain education program on concentration and prefrontal EEG activation of elementary students. Int. Brain Educ. Assoc. Acad. Semin. **2009**, 25–45 (2009)

6. Yu, S.J.: The effects of sensory motor program on change of electroencephalogram(EEG) and concentration of attention for mental retarded students. J. Korean Soc. Adap. Phys. Act. Exerc. **17**(1), 198–205 (2009)
7. Lokman, A.M.: Design & emotion: the kansei engineering methodology. Malays. J. Comput. **1**(1), 1–14 (2010)
8. Park, S.S., Hu, H., Lee, W.S.: A study on physiological signal changes due to distraction in simulated driving. J. Ergon. Soc. Korea **29**(1), 55–59 (2010)
9. Jee, S.D., Kim, C.B.: Objective evaluation of learning performance according to the color temperature of LED illumination. J. Korean Inst. Edu. Facil. **18**(2), 25–33 (2011)
10. Hwang, G.Y., Shin, D.K., Shin, D.I.: Design of biofeedback interface using biomedical signal analysis. Int. J. Inf. Technol. Comput. Sci. **7**(2), 104–113 (2012)
11. Jeong, K.I., Lee, C.W., Song, W.J., Sung, Y.J., Sun, J.S., Ryu, G.C.: Effects of correlated color temperature of LED light sources and illumination on visual performance at near. J. Korean Soc. Vis. Sci. **14**(3), 190–195 (2012)
12. Ko, J.K., Kim, J.H., Lee, M.J., Cho, M.R.: Evaluation and analysis of measurement data for lighting environment and user satisfaction in office. In: Proceeding of 2013 Spring Conference of Korean Institute of Illuminating and Electrical Installation Engineers, **2013** (5), 16–18 (2013)
13. ASHRAE Standard 55–2004 (2008) Thermal Environmental Conditions for Human Occupancy

The "Green" in Internet of Things: Case Study of Faculty Environment

Dimitar Ivanovski(✉)

Faculty of Electrical Engineering and Computer Science (FERI), University of Maribor (UM), Smetanova Ulica 17, 2000 Maribor, Slovenia
ivanovski.dimitar@gmail.com

Abstract. Building the internet of the future is on its way with the help of the Internet of Things. In order to benefit from the Internet of Things, first a smart environment has to be created. On the other hand, global trend of thinking is that we are moving towards a green future in which mankind will be more aware of its impact on the environment. This trend has started to significantly affect all of us, as well as companies, IT departments and faculties. This paper is a case study that discusses the energy efficiency of Internet of Things, and suggests introduction of an action plan for achieving sustain-ability objectives in a faculty environment. In the second and third chapter of the paper, the main questions of the case study are answered: Why care about the environment and what the advantages of Internet of Things are? Furthermore, in the fourth chapter, is described the way of connecting these two topics. In the fifth chapter, the plan and the systems that can be developed within the Internet of Things are presented. In the fifth chapter is presented the plan and the systems that can be developed within the Internet of Things. The paper concludes with the research challenges that hinder broader introduction of the concept of greener Internet of Things.

Keywords: Internet of things (IoT) · Environment · Devices · Green · Data

1 Introduction

The phrase "Internet of Things" was born as a title of a presentation that British engineer Kevin Ashton made at Procter & Gamble in 1999 [1]. Generally, it describes a unified identification of all physical things and their virtual presentation on the Internet [1, 2]. Today, Internet of Things (IoT) is no longer just title of science presentation – it is a science concept which connects the virtual world of information technology with the real world of things. IoT has a great potential and could touch many aspects of life — from home automation and healthcare, to infrastructure management, industrial applications and transport systems, to environmental monitoring and energy management.

And there are many advantages of integrating IoT into our lives. For example, alarm clocks can go off early if there is traffic on our way to work; medicine containers tells our family members if they forgot to take the medicine; plants communicate to the sprinkler system when it is time for watering; running shoes communicate time, speed

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 561–571, 2015.
DOI: 10.1007/978-3-319-21009-4_42

and distance so that their wearers can compete in real time with people on the other side of the world. The essence of the IoT, is that it will form a smart grid of “intelligent” devices and sensors in areas like smart homes, biomedicine, agriculture, mobile communications, transport and wearables.

In addition, the challenge of IoT is to provide new forms of communication between people and things, and between things themselves [3]. The IoT is a technological revolution that represents the future of information technology and telecommunications. Its development depends on the dynamic technical innovations in many important areas of wireless sensors, micro-electromechanical systems (MEMS) to nanotechnology and the Internet. Focal point of IoT is gathering data from multiple “things”. Here lies the great potential to use resources and things more efficiently [4].

On the other hand, the human population has been growing continuously, but the resources are finite, which means that if we use them continuously, we will eventually exhaust them [5]. In order to have a sustainable environment, people around the world will have to put a lot of effort to promote the importance of the reduction of carbon production as well as to draw attention to the benefits of reducing the consumption of energy.

The green potential of IoT is not only to enhance environmental sustainability but also to reduce costs by creating new opportunities [4]. The more devices and people connected, the better the opportunities. In a few years we will be communicating with a massive amount of devices, sensors and other physical and virtual “things”, which will also communicate with themselves, act “intelligently”, and produce green data for management and monitoring. The new things will also be independent and will be able to improve the quality of life and work in an efficient and optimized way [6].

Given the exploratory research questions, an interpretive case study approach was selected for this paper. Case studies can achieve a holistic understanding of cultural systems of action, and facilitate multi perspective analyses, encompassing not just the voice and perspective of the actors, but also those of the relevant groups of actors and the interaction between them. The interpretive case studies focus on the social construction of reality – how and why people see the world the way they do [7].

This paper is structured as follows. In the following section we discuss related environmental literature and the sustainable triangle. In section three, we explain the concept of IoT and look at the diffusion of IoT. In section four, we try to connect the two main topics of this paper: environment and IoT. In section five a faculty sustainable plan is presented. Then, in section six, the conclusions are drawn.

2 Environment and the Sustainable Triangle

The environment is undoubtedly changing and becoming warmer. Blinded by the lust for profit, today we live in a world where we are drowning in waste, breathing polluted air, eating poisonous food and sadly we keep destroying what is left. Most climate scientists agree that people are the main cause of global warming [8]. Taken from the article [9], China burns almost as much coal as the rest of the world combined. There should be no surprise that the global emissions are increasing again, after the fall during the economic downturn between 2008 and 2009 [10]. It is noteworthy that the IT

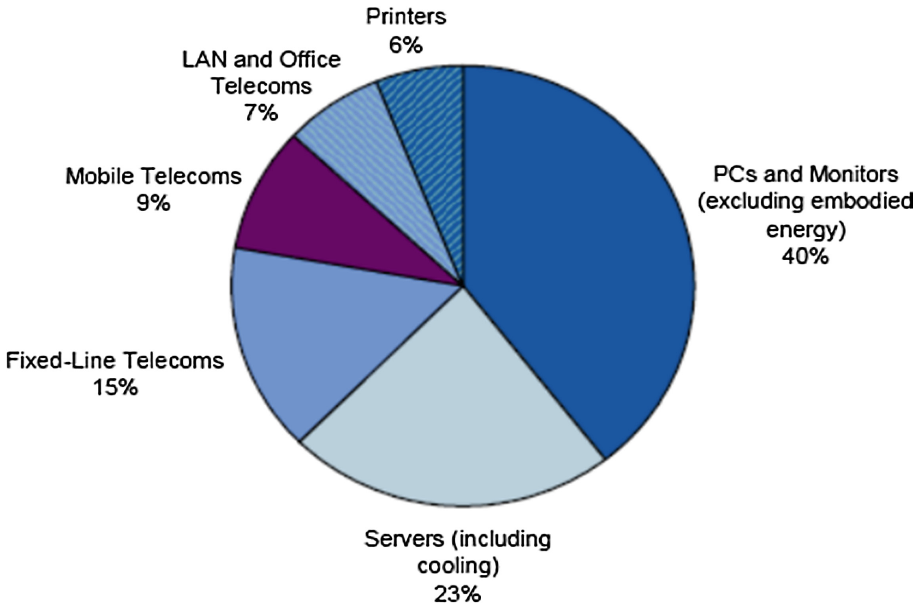


Fig. 1. Estimated distribution of global CO_2 emissions from ICTs [12]

contributes only 2 to 3 percent of greenhouse gas [11]. That includes emissions by ICT companies directly as well as energy consumption by ICT equipment. In Fig. 1, fixed-line telecommunications account for about 15 % of the total, while mobile telecommunications contribute an additional 9 % and LAN and office telecommunications about 7 % [12].

The vast majority of emissions usually come from sources that are not related to IT. According to the United Nations Environment Programme [5], building industry is responsible for more than 40 % of global energy use and one third of global greenhouse gas emissions, both in developed and developing countries. When we talk about IT and the environment, we need to think about the new role of IT, namely the creation of a more sustainable environment that offers economic benefits. This is the crucial time to look at it as a solution rather than a problem. IoT can be one of those solutions because it has great potential, is easy to upgrade or renew and can have significant environmental benefits [13].

While the causes and consequences for climate change are different, those causes should not be the crucial deciding factor when going "green". Perhaps the most significant one is to create values for the core of the organization. Green companies also represent a synonym for efficient operations. Well-known maxim is: "Business is IT". So we can say that green companies are "efficient IT". We will try to interpret this as an IT organizations attempt to achieve economic development and improve system performance [14].

Figure 2, shows that the IT affects business, which, in turn, influences the society and the overall environment in which the business exists.

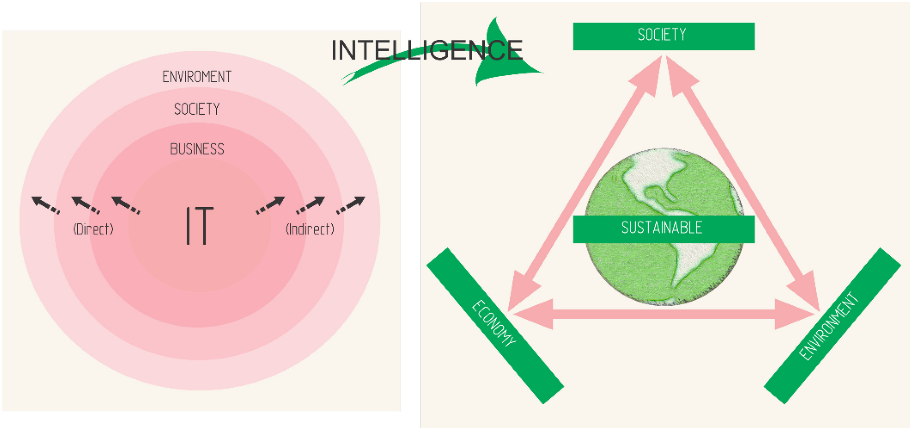


Fig. 2. IT impacts the business, society and the environment - leads to the sustainable triangle [14]

The left side of Fig. 2, shows that the IT is the core of the organization and it impacts on the business both directly and indirectly. Business then impacts on the society and the environment in which it exists. The direct influence of IT can be seen in the massive growth of household gadgets, use of computers at schools, faculties and hospitals, the high level of communication technology and the popularity of social networks.

A simple example of how IT affects the business, society and the environment is the Faculty of electrical engineering and computer science in Maribor (staff and students, as part of the campus) and the two biggest IT systems that the students and other employees are using: the Academic Information Subsystem (AIPS[1]) and the modular object-oriented dynamic learning environment (Moodle[2]). With the computerization of education we indirectly, and most probably, affected the number of jobs by replacing them with computers. Consequently, students of the faculty are not obliged to physically go to the location of the faculty to get academic information, but instead they can access everything from home. This includes downloading study and other academic materials; uploading homework and projects; taking quizzes from home; signing on and off for exams; and etc. Therefore, just by using AIPS and MOODLE they have social influence on the employees of the faculty and the environment. In addition, AIPS and MOODLE are using large data centers that consume a lot of electricity and generate waste products. This affects the whole environment, including the faculty itself where employees and students work and study. This can be considered as an indirect influence of AIPS and MOODLE on the society and the environment.

On the right side of Fig. 2, is the sustainable triangle that depicts the balance between society-economy, economy-environment and environment-society. If the

1 https://aips.um.si/.
2 https://estudij.um.si/.

Faculty of electrical engineering and computer science in Maribor wants to have strategic green IT initiative, then it has to answer these questions: How much can the faculty society financially bear? Is the environmental initiative practical? And is the environmental consideration fair and impartial? The fundamental principal of the sustainable triangle is to have the economic, social, and environmental factors in balance [14].

3 Internet of Things (IoT)

IoT is a new research in the field of Internet. It represents an advance version of Machine to Machine (M2M) Communication, where each object connects with another object, without human intervention [15]. Evolution of IoT enabled things/objects to get information about themselves, to interact with other things/objects, and to have access to data information gathered in their proximity. Development of IoT was also encouraged by the arrival of multisensory mobile devices such as smart phones. Current research expectation estimate that within 5–10 years, 100 billion devices will be connected to the internet [6]. Also, acquired from a Gartner article, the peak of inflated expectations for the IoT can be expected within 5–10 years [16].

It is said that regardless the statistics three things are clear [17]:

- The internet protocol IPv4 cannot satisfy the needs of so many connecting devices and the global adoption of IPv6 in the coming years will be critical for the successful development of the IoT in the future.
- Soon, multiple devices are going to be connected. Sensors that trigger lights, sensors for temperature, light and humidity, smart meters and others will be connected in a grid to remote datacenters and apps that can act accordingly to that data, save energy and respect the user preferences at the same time.
- Eventually the volume of communication traffic between devices (machine-to-machine, M2M) will exceed the volume of communication traffic between people (human-to-human, H2H) and between devices and people (machine-to-human, M2M).

Using the IoT has some advantages and disadvantages. The advantages are numerous and may far outweigh the drawbacks. There are advantages for health, safety, finance, and daily planning. For example, if appliances are able to communicate to each other, they can operate in an energy efficient manner. Also, relative advantages can be found in other fields of industry, environment and society. For example, in the social domain, there are opportunities to test the integration of different communities, cities, people, and especially services to improve the daily lives of people [18].

Compatibility is one the biggest problems for the IoT. That’s why there is a need for new standards that will enable products from different manufacturers to work together. Prominent international standardization bodies, such as IETF, IPSO Alliance and ETSI, are working on developing protocols, systems, architectures and frameworks to enable the IoT. At the moment it is hard to understand the scale of IoT because of its great complexity. We have to imagine that in the future of IoT all devices must work together and to be associated with all other devices and all devices must communicate seamlessly with related systems and infrastructures. The idea of the IoT is simple and interdisciplinary, but its implementation might be problematic [18].

At the moment, the acceptability for the IoT depends on the security and privacy of data that it provides. With the IoT a large amount of data will be collected, but afterwards someone will have to think about the security and how to protect the data. There is also the question who will control that data [18].

Regardless of the advantages and disadvantages, taken from the article [19], currently there is a strong placement of an IoT solutions among global enterprises. Even global companies recognize the transformative role of IoT [19].

4 The "Green" in Internet of Things

San Murugesan notes that Green IT may include: design, production, use and dispos-al of computers, servers, and associated subsystems such as monitors, printers, stor-age devices as well as networking and communication systems effectively and effi-ciently with minimal impact on the environment. In short, it is to reduce energy con-sumption and carbon footprint per IT-process. Also, we need to think about how, with the help of some IT devices, we can change the impact on the environment. San Murugesan also adds that "we are legally, ethically and socially responsible for the greening of our IT-products, applications, services and practices" [20].

The IoT will consist of a large number of devices and many of them will be embedded into the environment, and closely bound with physical/virtual things, such as some sensors. As shown in Fig. 2 the equipment of IoT can have direct and indirect impact. That impact may be power consuming or even electromagnetic pollution. So the possible influence on the environment exerted by the equipment of IoT may be more significant [21].

The bigger picture of Fig. 2 is the middle part that states that only with "intelligent use" of IT we can come closer to the sustainable triangle. An intelligent use can be the IoT which represents all the devices, objects and people connected in one network and gathering and processing data in order to lower the impact on the environment.

In the article [22], the author Hsing-I Wang suggests an architecture of green and smart campus (Fig. 3). The smart campus should represent a trendy application in the paradigm of the IoT. The benefits gained from building such campus will include the usage of all facilities to become more efficient and the energy consumed to be minimized. As stated in the article [22], "such efforts are also recognized as constructing a Green campus". The architecture is divided into the following three segments: hardware segment, the middleware segment and the presentation segment. All the data collected from the students, staff and building, via RFID, ZiggBee sensor modules etc., is collected and sent to the center of data and applications. The data is then computed, analyzed and controlled. The IoT is the connector of all the computers, emitters, receiver and participants in the green campus.

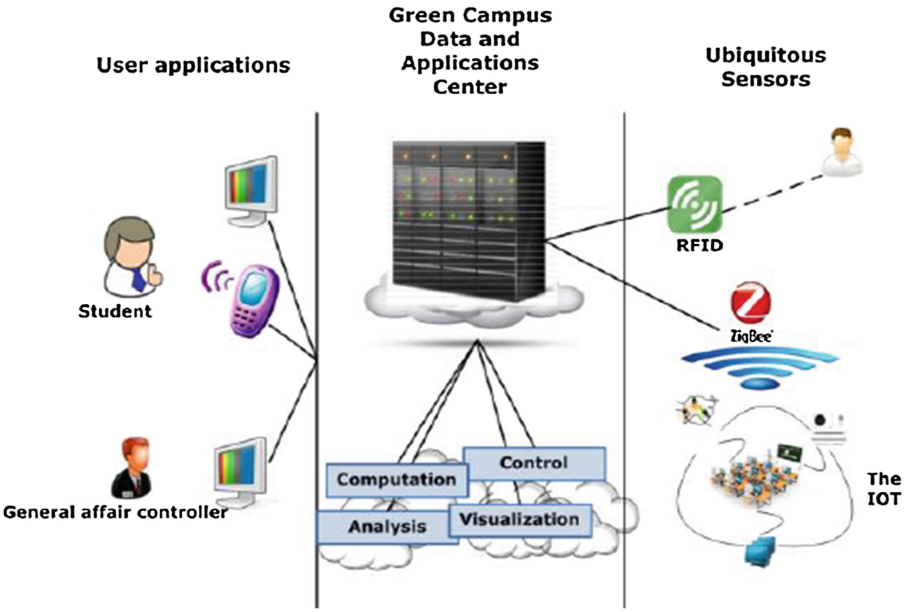


Fig. 3. Green campus architecture as proposed in the article [20]

5 Proposed Plan and Systems for Integration of IoT

In this chapter, we are going to try to answer the "how" question of this case study with a proposed plan for faculty environment. In that purpose, we have chosen the Faculty of Electrical Engineering and Computer Science at the University of Maribor (UM FERI[3]), which is one of the leading teaching and research institution in the field of Electrical Engineering and Computer Science in Slovenia. One of the key values for the faculty is social responsibility and sustainable development. By definition, sustainable development is defined as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs" [21]. In short, in order to meet the needs of the future we have to balance the social, economic, and environmental objectives. The same objectives are presented, and most importantly, are in line, with the objectives of the sustainable triangle (Fig. 2).

Furthermore, we are going to use PDCA (plan–do–check–act or plan–do–check–adjust) methodology, which is an iterative four-step management method used in business for the control and continuous improvement of processes and products [22]. PDCA is also known as a Deming circle. The methodology is an easy, four phase model for carrying out change in many different sectors, in a wide variety of situations. The first phase is to recognize opportunity and plan change. The second phase is to test the change and carry out a study. The third phase is to review the test, analyze the

[3] http://feri.um.si/.

results and identify what we have learned. The fourth phase is to take actions based on what we have learned in the study step [22]. We choose this methodology because its core is a continual improvement. After the fourth phase, if we are successful, we incorporate what we have learned from the tests into wider changes, and plan bigger improvements for the systems. (Figure 4)

1. Phase: PLAN.

IoT is an interdisciplinary field. On the other hand, at UM FERI there are 8 institutes and 30 laboratories. That's why there is a need of "common database of knowledge" or data and application center where all the researches, projects and student tasks are gathered and then presented in one responsive website so it can be accessible on multiple devices. The goal of this system should be knowledge management and connecting ideas from different institutes and laboratories, and creating a fertile ground for the IoT in the faculty. At the moment there is only MOODLE (as presented in chapter 2), but the big disadvantage of MOODLE is, that the sharing of ideas and knowledge is one directional – only from employees to students. That's why there is a need for a new, two directional platform, where students can share their ideas, knowledge and projects with every person in the faculty.

2. Phase: DO.

A significant part of testing the change is the building of the Faculty to become smarter. That means installing motion sensors that trigger lights, sensors for temperature, light and humidity, use smart meters that measures the energy consumption of the connected devices and finally create remote apps that can act according to that data, safe energy and respects the user preferences at the same time. Figure 5 shows an example of how a current lecture hall at UM FERI, can be transformed into a smart lecture hall.

3. Phase: CHECK.

The importance of monitoring and measurement of data is crucial. The faculty has to know how much it consumes so it can take the right corrective and preventive actions. All the records and data have to be collected and presented on a cloud based website. For that purpose, the faculty can use a similar architecture as shown on Fig. 3. The

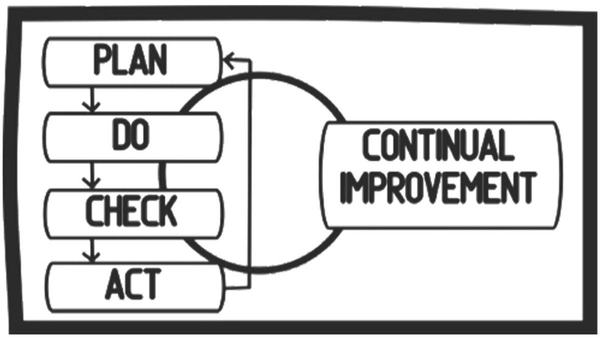


Fig. 4. Deming circle [22]

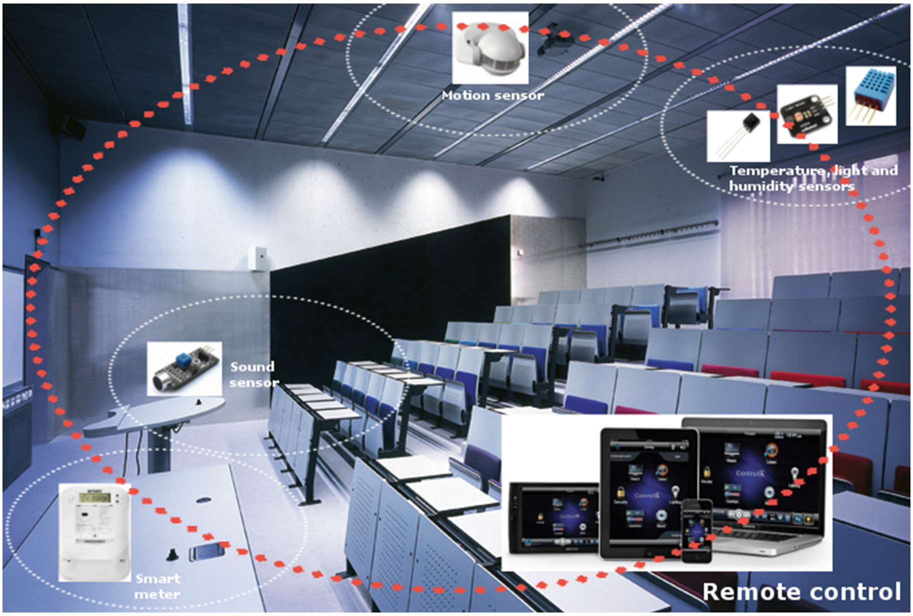


Fig. 5. Example of a smart lecture hall at UM FERI

website should also serve as visualization board, as well as controlling board for the management of the faculty.

4. Phase: ACT.

If results are not achieved, a new course of action has to be taken. The course of action should be about rectifying the failures of the devises and apps or their improvement. In this phase, it is very important to determine the profile of these actions.

The PDCA is assumed to be an ongoing process of performance improvement, with the cycle repeating until desired performance levels of the IoT, and impact on the environment is achieved.

6 Conclusion and Future Works

In this paper we presented a new approach to the concept of IoT – a greener one. IoT is a new research in the field of Internet, but with the right sustainable objectives, it can have significant environmental impact. Therefore, we can draw the following conclusions: Firstly, the environment is changing and we are responsible to have the economic, social, and environmental factors in balance. Secondly, the IoT is a field with a great potential that has interdisciplinary characteristics, and we have to take advantage of those. Thirdly, we gave a preliminary sight at the environmental impact of the IoT. Fourthly, we have shown how to plan for the IoT and what systems can be developed in a faculty environment.

Nevertheless, there is still a lot of work that need to be done. Future work involves in-depth investigation and planning of an IoT solution, apps and systems for a faculty environment that will have minimal impact on the environment. Also, we can look at the security issues of IoT which can focus on physical hardware security and privacy of data being collected and transmitted across the faculty network.

Acknowledgments. The author's gratitude goes to assistant Saša Kuhar for her advice, encouragement and support. This paper was also supported by the Faculty of Electrical Engineering and Computer Science at the University of Maribor.

References

1. Ashton, K.: That internet of things thing. RFID J. **10**(12), 1–2 (2009)
2. Borgia, E.: The Internet of things vision: key features, applications and open issues. Comput. Commun. **54**, 1–31 (2014)
3. Roman, R., Zhou, J., Lopez, J.: On the features and challenges of security and privacy in distributed internet of things. Comput. Netw. **57**(10), 2266–2279 (2013)
4. Li, L.: Application of the Internet of thing in green agricultural products supply chain management. In: 2011 Fourth International Conference on Intelligent Computation Technology and Automation, pp. 1022–1025, March 2011
5. United Nations Environmental Programme: Our Common Future: World Commission On Environment and Development. France, Paris (2009)
6. Valhouli, C.A., Group, T.H.: The Internet of things: Networked objects and smart devices, pp. 1–7. London (2010)
7. Walsham, G.: Interpretive case studies in IS research: nature and method. Eur. J. Inf. Syst. **4** (2), 74–81 (1995)
8. Reusswig, F.: History and future of the scientific consensus on anthropogenic global warming. Environ. Res. Lett. **8**(3), 031003 (2013)
9. Zhang, B., Chen, G.Q.: China's CH4 and CO2 emissions: bottom-up estimation and comparative analysis. Ecol. Ind. **47**, 112–122 (2014)
10. United Nations Environmental Programme (UNEP), "The Emissions Gap Report 2012," Nairobi, Kenya
11. Gartner, Gartner Estimates ICT Industry Accounts for 2 Percent of Global CO2 Emissions (2009). http://goo.gl/lEZq8. Accessed: 10 Jul 2013
12. Kumar, R., Mieritz, L.: Conceptualizing 'Green' IT and Data Center Power and Cooling Issues. Gartner, Research paper ID number G00150322 (2007)
13. Prasad, S.S., Kumar, C.: A green and reliable internet of things. Commun. Netw. **05**(01), 44–48 (2013)
14. Unhelkar, B.: Green IT Strategies and Applications: Using Environmental Intelligence, p. 480. CRC Press, Boca Raton (2011)
15. Kumar, C.: An energy efficient and reliable internet of things. In: 2012 International Conference on Communication, Information and Computing Technology (ICCICT), pp. 1–4 (2012)
16. Gartner, Gartner's 2014 Hype Cycle for Emerging Technologies Maps the Journey to Digital Business (2014). http://www.gartner.com/newsroom/id/2819918. Accessed: 07 Jan 2015
17. Mohor, M.: "Internet stvari – izzivi in priložnosti," Jožef Stefan Institute, Ljubljana, 1 (2014)

18. Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M.: Internet of Things (IoT): a vision, architectural elements, and future directions. Future Gener. Comput. Syst. **29**(7), 1645–1660 (2013)
19. Zebra Technologies’ Global Study Tracks the Growing Momentum of the Internet of Things in the Enterprise. https://www.zebra.com/us/en/about-zebra/media-center/press-releases/2014/forrester-2014.html. Accessed: 07 Jan 2015
20. Wang, H.: Toward a Green Campus with the Internet of Things–the Application of Lab Management. World Congr. Eng. **III**(5), 195–200 (2013)
21. Our Common Future: World Commission On Environment and Development. http://www.amazon.com/Our-Common-Future-Oxford-Paperbacks/dp/019282080X. Accessed: 04 Feb 2015
22. Micić, Ž., Micić, M., Blagojević, M.: ICT innovations at the platform of standardisation for knowledge quality in PDCA. Comput. Stand. Interfaces **36**(1), 231–243 (2013)

The Research on Path-Based DoS Attack Detection Algorithm in Wireless Sensor Networks

ZhenJiang Zhang[1(✉)], Dong Chen[1], Yun Liu[1], Fei Xiong[1], and Han-Chieh Chao[2]

[1] Key Laboratory of Communication and Information Systems, School of Electronic and Information Engineering, Beijing Municipal Commission of Education, Beijing Jiaotong University, Beijing 100044, China
zhangzhenjiang@bjtu.edu.cn
[2] Department of Electronic Engineering, National Ilan University, Yilan 26047, Taiwan

Abstract. This paper studies the problem of DoS attacks in wireless sensor networks (WSNs) security threat. Considering the Path-based DoS (PDoS) attack, this paper proposes a PDoS attack detection algorithm based on outlier mining in clustering (PDADBC). First, we give the definition of outliers under the background of PDoS attack of WSNs. The proposed PDADBC mainly include two periods, the first one is data pre-processing, in which the two-dimensional data are normalized to ensure the outlier detection in the following period more efficiency and accurate. After data pre-processing, the second period will perform the outlier detection algorithm, which is based on the improved DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm. The simulation results show that the proposed detection algorithm can effectively detect the PDoS attack.

Keywords: WSN · Dos attacks · Wireless sensor networks (WSNs) · Outlier mining · DBSCAN

1 Introduction

Wireless Sensor Networks (WSNs, wireless sensor networks), as one of the most important and most basic technology of Internet of things, has been widely applied in various fields because of its advantage of integration, miniaturization and network. The security issues of WSNs has attract much attention. Since WSNs consists of a large number of wireless sensor nodes, the nodes are energy limited and capacity limited [1], and the characteristics make the security of the network challenged, it is vulnerable to various attacks. In all of the attacks, denial of service (DoS) attack is the most frequent one and it has many kinds of attacks.

Deng et al. in [2] proposes a path based DoS (Path-based DoS, PDos) attack on application layer. This attack exists in the special network structure, requires the sensor network is the two level hierarchical structure with the cluster-head nodes and the nodes in cluster. At the same time, the network also has the dedicated path nodes to

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 572–587, 2015.
DOI: 10.1007/978-3-319-21009-4_43

forward and transfer data. The attacker first will capture the cluster-head node or the nodes in cluster, then arrive at the station by the way of multi hop through their legitimate large data packet flooding along the transmission path. The purpose is to occupy the network bandwidth, consume the energy of the path nodes, and lead to the death of the path nodes, and then make the network paralyzed. Therefore, the most fundamental purpose of PDoS attack is to deplete the energy of the core path node. If an attacker can move at any time, so it can simultaneously capture more than one cluster-head nodes, and then launches distributed DoS attack. But usually, the attacker is fixed and captures the closest node. Normally, PDoS attack is a centralized attack. And this article also only considers this attack mode.

In order to resist PDoS attack, path node should detect the data packets at first, and then discard the attack packets. In the literature, researchers adopt two kinds of conventional methods to resist attacks, one is to establish a shared key pair mechanism between the source node and path node and resist the attacks by means of encryption; the other is the rate control which constrains the number of data packets transmitted by the path node for per unit time. However, due to the harsh requirements of packet size in WSNs and different network traffic of different nodes, the two ways exist inherent drawbacks, and are not entirely suitable for WSNs environment. Therefore, based on the basic thoughts of the two methods, gradually appeared some new mechanisms to resist attacks. But it also need path node verify the authenticity of data packets, and then decide whether to forward or discard, which undoubtedly increases the communication overhead.

Aiming at the problem that when path node tests the authenticity of the data packet, it will consume energy to a great extent, this paper proposes that we do not detect the attack in the path node, but utilize the advantages of base station in energy, calculation ability, memory and communication ability to analyze the uniqueness reflected on the base station when the network suffers the PDoS attack, carrying out the detection algorithm at the base station, detecting the PDoS attack.

This paper is structured as follows: Sect. 2 introduces the existing settlement mechanism of PDos attack and its advantages and disadvantages; the DBSCAN algorithm and its inherent defects are described in Sect. 3; In the Sect. 4, this paper proposes the PDoS attack detection algorithm based on improved DBSCAN; Sect. 5 carries on the performance analysis and evaluation of the improved DBSCAN algorithm by using the simulation software; finally Sect. 6 summarizes the contribution of this paper and make the forecast to the future research direction.

2 Related Work

In this section, we review some related work on our research in the field of DoS attacks in WSN.

Anthony Wood and John Stankovic et al. first summarized DoS attackmodes in WSNs in 2002 [3]. David R and Raymond are summarized and perfected threat and response measures of WSNs in the presence of the time in 2008 [4]. The classification of DoS attack and defense measures in WSNs includes DoS attacks in physical layer, Link layer, Network layer, transport layer and application layer. Jamming attack is the main form of attack aimed at a physical layer in WSNs. Wenyuan Xu and colleagues proposed a mechanism of jamming attack detection in WSNs, and the attack is classified as: continuous type, confusion, random type, type reaction [5]. TinyOS Destination-Sequenced Distance-Vector Routing using the bidirectional link of high quality constitute the propagation path, and increase the link quality measurement device, can realize the mode of defending jamming attack [6]. The denial-of-sleep attack is a common DoS attack in data link layer, it will prevent the RF module of node into sleep mode [7]. Frank Stejano and Ross Anderson first proposed this kind of attack, and this form of attack named sleep deprivation [8]. At the same time, the author studied the effect of this form of attack on the mobile device battery energy. Garth Crosby, Niki Pissinou and James Gadze proposed a security protocol based on cluster head selection of trusted framework in the ad hoc network [9]. This agreement is very effective in resisting the cluster attacks, but it requires large bandwidth and large clusters, and the bidirectional encryption, security mechanism is not entirely suitable for WSNs.

Since using encryption and rate control exist congenital defects, scholars around the world are in reference to the basic ideas on the basis of the two methods and they put forward some new mechanisms of resistance to attack.

Deng et al. proposed a lightweight one-way hash chain security authentication mechanism [9, 10]. This mechanism is to let each sending node joint its path node establishes a hash chain by using one-way hash function and adds a hash value to each of the sending data packet. When the data packet arrives at a path node, path node calibrates the hash value in the packet. The packet which is success in the check continues to be forwarded to the next path node. Continue executing the inspection until the packet arrives at the base station. If the check failure to a certain number, it can be certified that the data packet is an attack packet and the path node will throw away the packet. This mechanism is a lightweight authentication mechanism. But this mechanism due to need increase hash value in every sending packet and perform hash check in each path nodes, it increases energy consumption at the same time also increases the time of transmission the data packet. What's more, if the attack node capture the source node, then it masters the whole hash chain, this mechanism becomes a complete failure at this time.

A. Perrig et al. proposed a reliable broadcast protocol μTESLA (Timed Efficient Stream Loss-tolerant Authentication) under the WSNs environment. This agreement is adding a message authentication code (MAC) in the broadcast messages by one-way hash chain. First, hash value will be allocated in each time slot and the hash value is used to produce the MAC value in this slot of sending data packets. In order to solve the problem of the loss of data packets, μTESLA agreement introduced a multi-level one-way hash chain, using the secondary OHC to start the first level of OHC [11].

Y. Hu et al. proposed a security routing protocol according to the demand type in WSNs and introduced OHC mechanism to solute the flooding of the route request information. When sending node broadcast the path request information, it will add OHC value in the information and other nodes will perform broadcast information check by checking OHC value. OHCs are also adopted by the INSENS mechanism to withstand the broadcast flooding of the path update information.

Recently, en-route filtering schemes is widely used to solve the PdoS attack in WSNs. The core idea of this mechanism is to share the key pair between the path nodes and the cluster head nodes. The cluster head node uses the shared key to generate message authentication code(MACs, and path node forwards data packet after checks it. In the mechanism of SEF, the Bloom filter can reduce the length of the MACs while ensuring safety, and path nodes and the cluster head nodes use pre built key generating message authentication code. In this mechanism, false packets will be discarded by path nodes at any time. However, this mechanism also has some problems. The SEF mechanism uses a probabilistic algorithm. That is, it can not guarantee that every false data packet can be filtered out. In addition, the overhead of the SEF system is relatively large.

Zhu et al. proposed a kind of coping mechanism of cross keys [12]. The cluster head node and path node build the cross key according to the random prior distribution key. These keys and the hop-by-hop authentication guarantee the captured nodes can be detected by the base station when their number exceeds a certain threshold. The problem of the mechanism is: there is no authentication mechanism which is effective and suitable for the multi hop between two nodes, and the establishment of key pairs between multi hop nodes need for large communication overhead, and the processing speed is slow.

The summary can be found, these technologies which coping PDoS attack no matter what mechanisms they adopted, are needed the path nodes to verify the authenticity of the data packets, and then decide whether to continue to forward or discard. Whether it is a one-way hash chain, message authentication code, or use the key pair will undoubtedly have greatly increased the cost of communication, these factors will eventually consume a great extent energy. Consequently, this paper will not conduct the attack detection in the path node, the centralized attack PDoS detection method performed in base station will be adopted (Fig.1).

3 Preliminaries

3.1 The Definition of PDoS Attack

Deng et al. proposed DoS(Path-based DoS) attack based on the path of application layer in 2005. This attack exists in the network structure more special, require sensor network has two level hierarchical structure, with clusters and nodes. At the same time, the network also has a dedicated path nodes to forward and transfer. The attacker first will capture the head nodes or nodes in the cluster, and then through their legitimate large data packet flooding along the transmission path by way of multi hop arrived at the base station, the purpose is to occupy the network bandwidth consumption path, the energy of the node, and then leads to the death of path nodes, so that the whole network or partial network into paralysis. Then through them adopt the mode of flooding along the transmission path send large of legitimate data packets by way of multi hop to arrived at the station, The purpose is to occupy the network bandwidth, consump the energy of path nodes, and lead to the path nodes death, so that the whole network or partial network into paralysis. Because in WSNs network based on path, the death of upstream path node will lead to the transmission data packets of downstream sensing nodes and the cluster head node is blocked, so the network had to be paralyzed. Therefore, the most fundamental goal of PDoS attack is to deplete the core path nodes' energy. If an attacker can move at any time, so it can simultaneously capture more multiple cluster head nodes, and then launch distributed DoS attack. But usually, the attacker is fixed, and capture the nearest nodes. Thus, PDoS attack is normally centralized attack, this paper only considers the attack mode. The structure of the network under PDoS attacks as shown in Figs. 2–3.

3.2 Definition PDoS Attack Outliers

In WSNs, PDoS attack node attacks in the form of sending a large number of data packets. The purpose is to consume the energy of the path of nodes as quickly as possible, so that the network paralysis. Therefore, abnormal data from attackers triggered a series of abnormal behavior is so called PDoS attack outliers in WSNs.

In order to achieve the goal node as soon as possible, the number of packets attack node sends is far greater than other normal nodes, while its energy consumption will be far greater than other normal nodes. Therefore, the two indicators of the number of data packets the node sends and the energy consumption values form a two-dimensional data in this article, so, the two-dimensional data of the attacker node is the data space outlier.

Hypothesis: a sensor network consists of N nodes. The number of packets node i $(1 \leq i \leq n)$ sends in a certain period of time is denoted by P_i. The energy consumption of this time period represented by E_i.

The two-dimensional data of the i-th node is (P_i, E_i). Data space composed of two-dimensional data for each node is $\Theta = \{(P_1, E_1), (P_2, E_2), \ldots\ldots, (P_n, E_n)\}$.

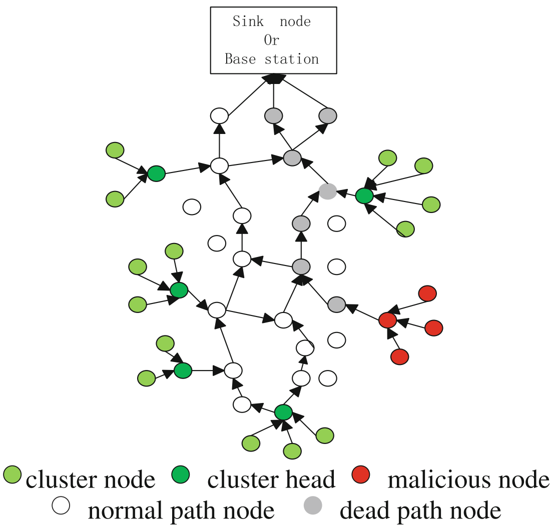


Fig. 1. WSNs attacked by PDoS

In particular, although the number of packets and the energy consumption of these two data are relevant, and by reducing the dimension, two-dimensional data can be turned into a one-dimensional data. Since the path node occurs in certain circumstances will send the same data packet repeatedly, when the two related data is no longer absolutely. Thus, the definition of two-dimensional data space has real meaning and value. As the traditional PDos attack algorithm of detecting the end of the path nodes requires a lot of communication overhead, it's not available for WSNs. Therefore, in order to detect the abnormal data (outliers) in the two-dimensional data space effectively, in this paper, we use the advantages of centralized base station to propose a detection algorithm based on clustering of outliers.

3.3 DBSCAN (Density-Based Spatial Clustering of Applications With Noise) Algorithm

Currently, the main clustering methods can be divided into the following categories: (1) hierarchical clustering method. (2) grid method. (3) density clustering method. (4) the division clustering method. The core idea of the density -based method is, if the density of the data object in a zone exceeds a preset threshold, it is added to the closest class. Density -based clustering method is different from the algorithm that based on a variety of distances for data set processing clustering. It is based on the density of the data set. Since in the WSN, the distribution of the data set made up of data package and node energy is not necessarily uniform, therefore, this paper applies DBSCAN algorithm of density-based clustering method for detection of outliers.

DBSCAN is a density-based clustering method based on high-density area of connectivity, in which cluster is defined the largest collection of connectivity points. It is mainly used two parameters: radius *eps* and density threshold. *MinPts.* DBSCAN algorithm requires a subjective selection of these two parameters, and the selection of parameters determines the outcome of the final clustering. In terms of computational complexity, if DBSCAN use spatial index, the computational complexity is $O(n\log(n))$, otherwise, it is $O(n^2)$.

There are some inherent flaws in DBSCAN algorithm such as sensitive parameters 、not ideal clustering effect for data space of non-uniform density 、high I /O overhead and high computational complexity and so on. This paper improves DBSCAN for its sensitive parameters and non-uniform distribution density and not ideal clustering effect. And this paper increases the effectiveness of the clustering while improving the accuracy of detection of outliers when PDoS attacks.

4 PDoS Attack Detection Algorithm Based on Improved DBSCAN (PDADBC)

The defect of DBSCAN algorithm is mainly due to parameter radius and density threshold. In this paper, we use K-nearest neighbor based on the DBSCAN algorithm, prejudge the density distribution of the data points and adjust the radius value dynamically to instead of the value set by the prior knowledge in advance, while solving the uneven distribution of density and clustering effect problem.

4.1 Sample Data Preprocessing

In order to make the detection process more efficient, more accurate detection results, this paper first for two-dimensional data has been sampled data preprocessing.

Suppose that there are n nodes, two-dimensional data of each node can be expressed as:

$$n_1 : (P_1, E_1), n_2 : (P_2, E_2), \cdots\cdots, n_i : (P_i, E_i), \cdots\cdots, n_n : (P_n, E_n)$$

The data space need to detect: $\Theta = \{(P_1, E_1), (P_2, E_2), \cdots\cdots, (P_n, E_n)\}$

First of all, through the formulas 1 and 2 were obtained by the number of data packets and energy consumption expectations of known data space:

$$\overline{P} = \frac{P_1 + P_2 + \ldots + P_n}{n} \tag{1}$$

$$\overline{E} = \frac{E_1 + E_2 + \ldots + E_n}{n} \tag{2}$$

Then, make use of the formulas 3 and 4 are obtained the variance of the number of data packets and energy consumption of each node:

$$S_{P_i}^2 = \left(P_i - \overline{P}\right)^2 \tag{3}$$

$$S_{E_i}^2 = \left(E_i - \overline{E}\right)^2 \tag{4}$$

So, each node after two-dimensional data preprocessed can expressed as:

$$n_1 : \left(S_{p_1}^2, S_{E_1}^2\right), n_2 : \left(S_{p_2}^2, S_{E_2}^2\right), \cdots\cdots, n_i : \left(S_{p_i}^2, S_{E_i}^2\right), \cdots\cdots, n_n : \left(S_{p_n}^2, S_{E_n}^2\right) \tag{5}$$

The data space into $:\Theta = \left\{\left(S_{p_1}^2, S_{E_1}^2\right), \left(S_{p_2}^2, S_{E_2}^2\right), \ldots\ldots, \left(S_{p_i}^2, S_{E_i}^2\right), \ldots\ldots, \left(S_{p_n}^2, S_{E_n}^2\right)\right\}$

Through the variance is calculated for each data, then by the two-dimensional variance data component to be detected data space, data pretreatment such without changing the data set characteristics under abnormal data and normal data between the distance to pull big, can make the abnormal data in the next step of detection is more efficient and accurate.

4.2 Parameter Setting

a. The definition of density
Definition 8: density: For a given object set $D, \forall p \in D$. The point P is define as reciprocal average distance density of the point P and k nearest neighbors of it. The calculation formula is as follows:

$$density(p) = 1/\left[\frac{1}{k}\sum_{i=1}^{k} dis\tan ce(p,i)\right] \tag{6}$$

Where $dis\tan ce(p,i)$ the point and the nearest neighbor distance function, density is reflects the point P in the vicinity of the sparse distribution of the data level. Here by the Euclidean distance as the distance metric, the mean distance decreases, the more dense the points.

b. Setting the value of radius *eps*

The value of radius *eps* for the DBSCAN algorithm needs for subjective setting based on the prior knowledge, leading to some problems such as choosing parameters difficultly, sensitive parameters and the poor cluster effect resulting of uneven distribution. To solve the problem, in this paper, we set and dynamically adjust the value *eps*, based on the k-nearest average distance and the degree of distance deviation.

Firstly, we calculate k-nearest average distance $dis\tan ce(\bar{p})$about point p according to the formula 7;

$$density(\bar{p}) = \frac{1}{k}\sum_{i=1}^{k} dis\tan ce(p,i) \tag{7}$$

Then, we calculate the degree of distance deviation between k-nearest average distance and the distance which is between core object p and each k-nearest node, based on the formula 8:

$$dis\tan ce(s) = \sqrt{\frac{(dis\tan ce(p,1) - dis\tan ce(\bar{p}))^2 + \ldots + (dis\tan ce(p,k) - dis\tan ce(\bar{p}))^2}{k}} \tag{8}$$

So, the value of *eps* is:

$$eps = dis\tan ce(\bar{P}) + dis\tan ce(s) \tag{9}$$

The setting of k-nearest average distance $dis\tan ce(\bar{P})$ and the degree of distance deviation $dis\tan ce(s)$ about the core point p makes the algorithm both not define the normal data as outliers because of too low value of *eps* and add outliers to the normal data cluster because of too high value of *eps*.

The sum of k-nearest average distance and the degree of distance deviation determines that the value of *eps* is not too low, and removes one or more points which is furthest from others and makes a great effect on the value of *eps*, making the value of *eps* be setting reasonably according to the density of core point p.

Although the value of k selected still impacts on the value of *eps*, thereby affecting the results of clustering, the selection of the value k is more rational to be found. Moreover, when we select value k, the sensitivity of the clustering results is clearly much lower than the method which sets directly the value *eps*.

Now, the setting of parameter for DBSCAN: (1)*eps*, the search radius (2)*Minpts*, threshold density, is converted into the present parameter: (1)p, the number of neighbors about object p (2)*Minpts*, threshold density.

4.3 Description on the Implementation of the Algorithm

In the Sect. 3.2 of this paper, we have defined sample data space and outliers under the PDoS attack in WSNs, that is, the collection which is two-dimensional data consisting of the number of packets transmitted by each node and the energy consumption of nodes is the sample data space needing to be detected $\Theta = \{(P_1, E_1), (P_2, E_2), \cdots\cdots, (P_n, E_n)\}$; and the outliers is the abnormal 2-dimensional data points generated by attack nodes in the sample space.

The flow chart about implementation on algorithm shows as Fig. 2.

Outlier detection is mainly divided into the following processes:

1. We preprocess the sample data space to be detected by the Eqs. (1)–(4). We get variance of data collected directly $\Theta = \{(P_1, E_1), (P_2, E_2), \cdots\cdots, (P_n, E_n)\}$, and invert them into $\Omega = \left\{\left(S^2_{p_1}, S^2_{E_1}\right), \left(S^2_{p_2}, S^2_{E_2}\right), \ldots\ldots, \left(S^2_{p_i}, S^2_{E_i}\right), \ldots\ldots, \left(S^2_{p_n}, S^2_{E_n}\right)\right\}$, increasing the distance between the data. The preprocessing of data, under the premise of not affecting the data itself characters, widens the distance between the normal and abnormal data, facilitating the next step for abnormal data detection.
2. Setting the parameters K and *MinPts*. The selection on parameter K judged by the node deployment, deployment density and distribution of data-aware, *MinPts* < *k*;
3. Firstly, according to the formula (6), we can calculate the density of each data object *density (p)* about data space, build a density list of data space sorted according to density, and mark the minimum value of the density list corresponding to the data object as suspected outliers. For example, if the density list arranges in descending order: *density* (p_1) > *density* (p_2) > ... > *density* (p_{10}), point p_{10} is the point of the minimum density, we mark the point p_{10}.
4. Selection on the point with the highest current density and not been marked from the density list as the core node;
5. Calculate the *eps* value of the core node radius based on formulas (7)–(9), and check on the neighborhood of the core node. If the number of contained objects is no less than *MinPts*, we will create a new cluster C, adding all neighborhood nodes to C.
6. Checking the neighborhood of data object q that has not been processed in C, if the data object contained is more than *MinPts*, we add objects which do not belong to any one cluster to C.

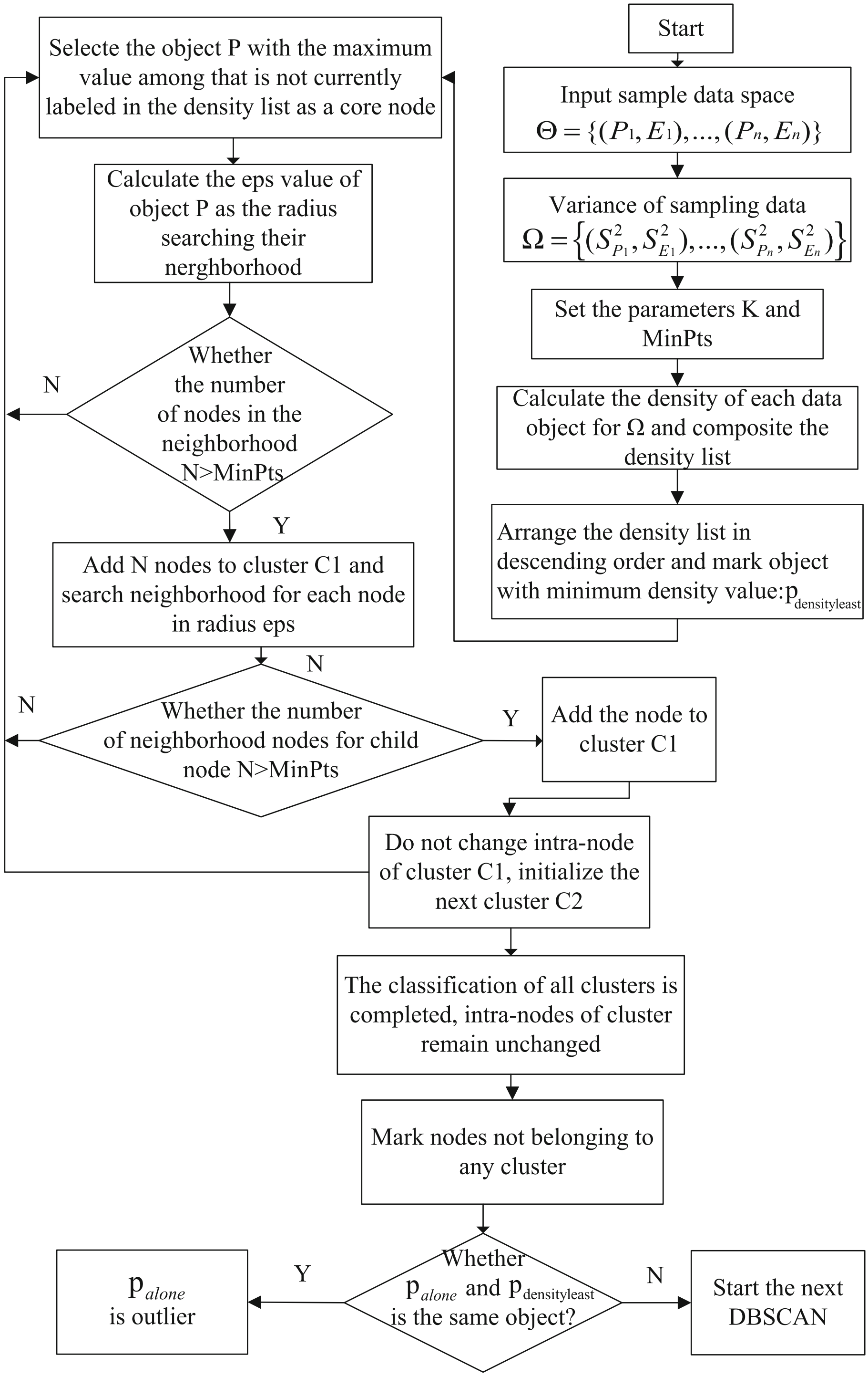


Fig. 2. Flow chart of PDoS attacking detection algorithm based on outliers mining

7. Repeat step 6, continue to check unprocessed object in C until no new objects added to the current cluster C;
8. Repeat steps 4-7 until all objects of the data space form stable clusters.
9. Mark data objects not belonging to any clusters as suspected outliers.
10. Comparing suspected outliers gained from step 3 with that gained from step 9, if they belong to the same data object, we thick the object as the outlier.

Algorithm 3.1. Outlier detection algorithm based on the improved DBSCAN algorithm

input：the set of data objects Θ, the number of neighbors *k*, density threshold *MinPts*, *MinPts* < *k*

output：outlier

DBSCAN（Θ, *k*, *MinPts*）

1: Begin

2: Convert Θ into Ω;

3: $density(p)=\frac{1}{k}\sum_{i=1}^{k} distance(p,i)$; //calculate the density value of each data object for data space;

4: rank the density(p) list, and mark the minimum object: $P_{densityleast}$; //arrange the density list in descending order and mark object with minimum density value;

5: For each p, p has the current maximum in the density list; //select the data object with the maximum density value and not been marked from the density list;

6: mark p as the core point and visited; //mark data objects

7: C = next cluster, init C=0; //the number of initialized clusters is 0;

8: $eps = dis\tan ce(\overline{p}) + dis\tan ce(s)$; //calculate the radius of point p *eps*;

9: N = getNeighbours (p, *eps*); //

10: if sizeof(N) <

11: C = next cluster; //create a new cluster C;

12: Else add p to cluster C; //add all point of p to C;

13: for each point p' in N //cycle all points in the neighborhood p;

14: N' = getNeighbours (p',); //conduct radius check on all point in the neighborhood p';

15: if sizeof(N') >= *MinPts* then

16: N = N+N'; //if it is more than *MinPts*, extend the number of N;

17: end if

18: end for

19: end for

20: mark the object ,who don't belong to any cluster; //mark nodes not belonging to any cluster;

21: P_{alon}e &&; //compare two nodes whether belonging to the same data object;

22: if P_{alone} is the

23: then P_{alone} or $P_{densityleast}$ is the outlier; //if so, it is the outlier;

24: else P_{alone} or $P_{densityleast}$ is not the outlier;

25: begin next round DBSCAN //if not, start the next test;

26: End

Fig. 3. The pseudo-code of outlier detection algorithm based on the improved DBSCAN algorithm.

5 Simulation Results

In order to verify the effect that we adopt the proposed improved DBSCAN algorithm to detect the outliers, this section uses the simulation tool MATLAB to test the validity of the algorithm, and compares the test results with the DBSCAN algorithm. The data set used in the procedures has 100 discrete points, and the data distribution is basically in accordance with the data set defined by PDoS outlier in Sect. 3.2, which is shown in Fig. 4. The abscissa of the testing data set D denotes the number of data packets transmitted by the node, and the ordinate denotes the node's energy consumption. In order to facilitate the analysis and description, through the observation of data set D, this paper names the node cluster in lower left quarter "Cluster 1" and names the middle big cluster "Cluster 2".

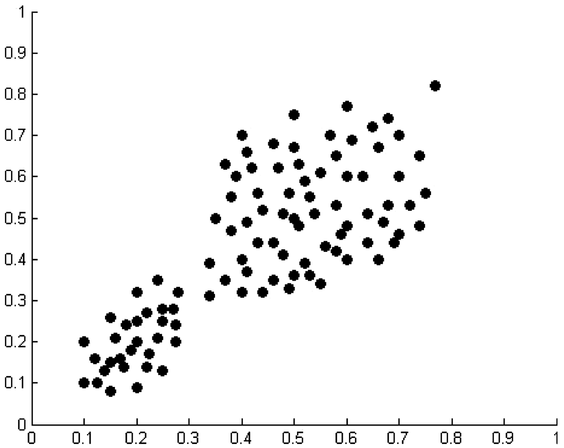


Fig. 4. The set of testing data D

1. Using DBSCAN algorithm for outlier detection of data set D
   As mentioned above, DBSCAN algorithm has two core parameters: radius *esp* and density thresholds *MinPts*, which need to be set according to the prior knowledge, and they have high sensitivity. The parameter value is too large or too small will lead to inaccurate clustering results. Aiming at the outlier detection, the detection results of the algorithm is that the node is outlier but not be detected, or the node is not outlier but is defined as outlier. Based on this, we set up two sets of parameter values, the result is compared as follows:

   (1) $esp = 0.05$, $MinPts = 5$;
   The operating result of data set D through the DBSCAN algorithm with the two parameters is shown in Fig. 5. The result shows that three of the original "Cluster 1" are out of the cluster, becoming outliers; four of the original "Cluster 2" are also out of the cluster, becoming outliers; nodes in the upper right corner are becoming outliers. The reason is that the value of *esp* is too small and the value of *MinPts* is too large. Therefore, in the parameter setting, if *esp* is too small but *MinPts* is too large, each node of the data set will be defined as a outlier.

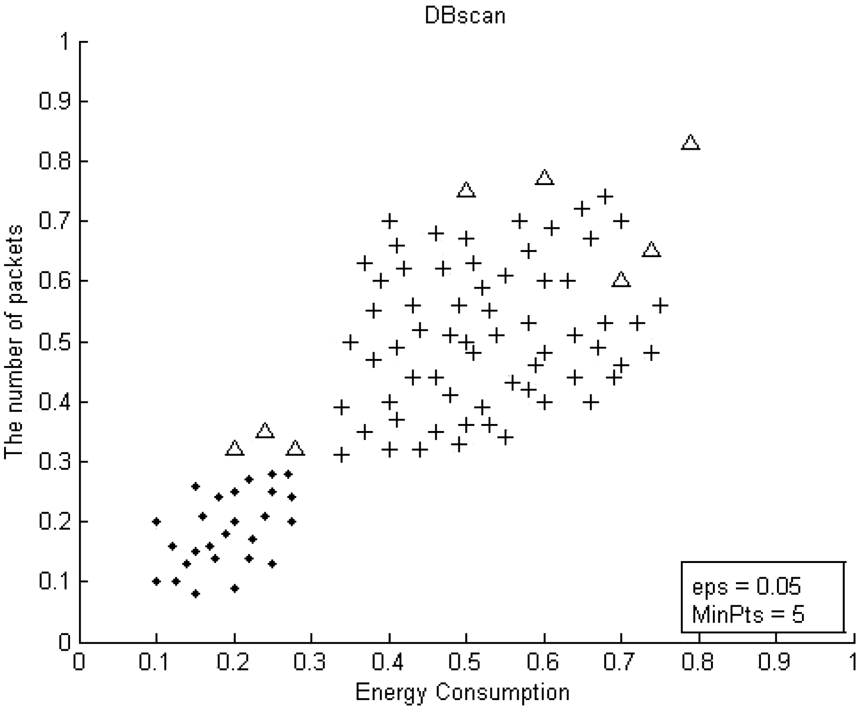


Fig. 5. The clustering result of data set D by DBSCAN (1)

(2) *esp* = 0.15, *MinPts* = 2;
The operating result of data set D through the DBSCAN algorithm with the two parameters is shown in Fig. 6. The result shows that the two clusters of the original data set, "Cluster 1" and "Cluster 2", as well as the independent nodes in the upper right corner, are classified together, becoming a new large cluster. The reason is that the value of *esp* is too large and the value of *MinPts* is relatively too small. Therefore, in the parameter setting, if *esp* is too large but *MinPts* is too small, the data set data will not be clustered, and the outliers will not exist.

Therefore, it is not easy to set the values of search radius *esp* and density threshold *MinPts* in the DBSCAN algorithm subjectively, and the detection result is not ideal.

2. Using the improved DBSCAN algorithm for outlier detection of data set D
In the improved DBSCAN algorithm, we no longer set the value of *esp* directly, but introduce the *k* nearest neighbor, judging the density distribution around the data point, dynamically calculating the value of search radius *esp*, instead of the *esp* value which is set in advance according to the prior knowledge. The dynamic calculation values are not too large, nor too small, which avoids the disadvantage caused by the too large and too small values. Although the selection of *k* value will

still affect the value of *esp*, and then influence the clustering results, but the *k* value selection is more rational. Because the *k* value is set mainly depends on the size and distribution of the data set, and the two indexes are more intuitive and easier to get. More importantly, the value of *esp* is dynamically calculated by the *k* value, so set the value of *k* firstly, the sensitivity of the clustering results is significantly lower than the directly setting of the value of *esp*.

(1) $k = 3$, $MinPts = 2$;
The operating result of data set D through the improved DBSCAN algorithm with the two parameters is shown in Fig. 7. The result is shown that the algorithm classifies the two clusters, at the same time independent node in the upper right corner is detected, because this node does not belong to any cluster, it is defined as the outlier. Outlier is defined that the number of data packets and node energy consumption form the two dimensional data. Under this definition, the attacking node data reflects in the two-dimensional space which distributed in the upper right side of the data collection in this coordinate system, which is consistent with the detection algorithm. Therefore, the algorithm has a high degree of consistency in theoretical analysis and practical test, and the algorithm is feasible and effective.

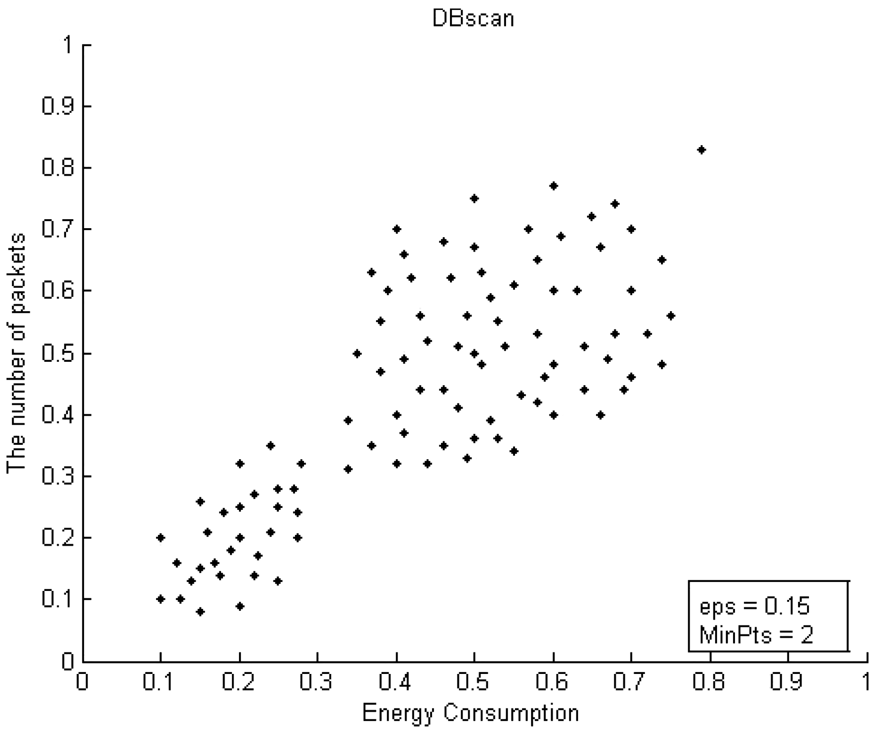


Fig. 6. The clustering result of data set D by DBSCAN (2)

$eps = 0.05$, $MinPts = 5$

$eps = 0.15$, $MinPts = 2$

(2) $k = 4.5$, $MinPts = 3,5$;
Re-setting the two sets of parameter values, running the improved DBSCAN algorithm, the detection results are not changing, which illustrates that the proposed algorithm reduces the sensitivity to parameter values, as long as the the value of parameter is in a range generally, the test results will not change. Therefore, the improved DBSCAN algorithm is better than the original DBSCAN algorithm in the detection accuracy and sensitivity to parameters.

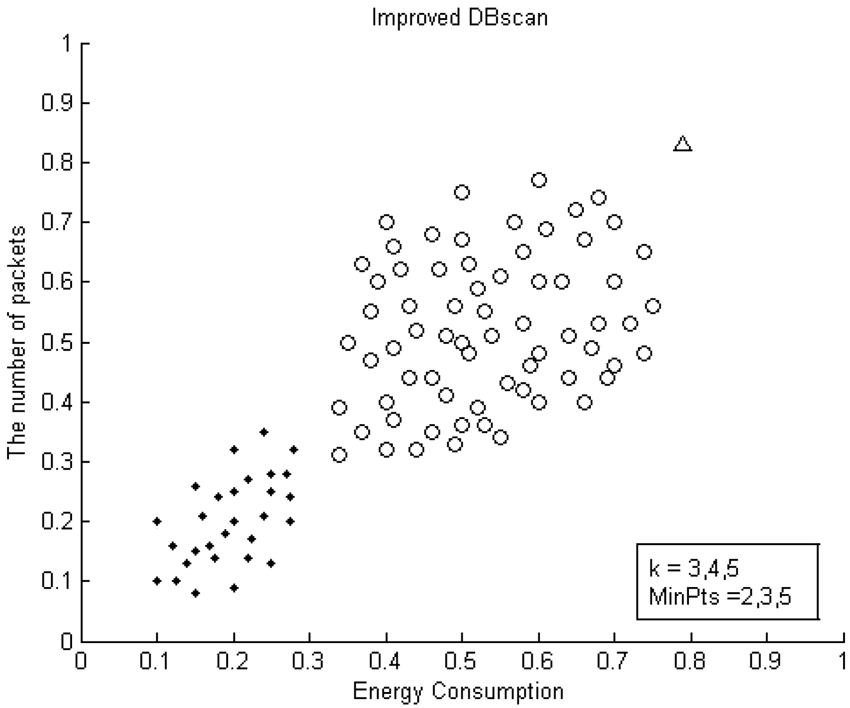


Fig. 7. The clustering result of data set D by Improved DBSCAN

6 Conclusion

This paper mainly studies the DoS attacks in wireless sensor networks (WSNs), detecting the PDoS attacks by using the thought of outlier mining. First of all, through analyzing the characteristics of PDoS attacks, this paper clears the definition of outlier under the PDoS attack behavior in WSNs. Based on this, this paper detects the outlier based on clustering, and improve and optimize the parameter value settings mode based on the DBSCAN algorithm which is a clustering method based on density. At the same time, in order to improve the accuracy of the detection algorithm, this paper presents sample data preprocessing, i.e., in the premise of not changing the data characteristics, using the idea of variance to increase the distance between data nodes in the data space, improving the clustering accuracy of detection. The simulation and the results analysis of the algorithm also prove that using outlier detection, mining the outlier points based on the clustering method of density is effective and accurate of the detection of PDoS Attacks behavior in WSNs.

In the future, we will focus on DoS attacks including but not limited to the PDoS attack in WSNs. Furthermore, DoS attacks in different layers such as physical layer, Link layer and Network layer would be researched.

References

1. Yulong, S., Qingqi, P.: Introduction of the security technology in wireless sensor networks. People's Posts and Telecommunications Press, Beijing (2008)
2. Deng, J., Han, R., Mishra, S.: Defending against path-based DoS attacks in wireless sensor networks. In: Proceedings of 3rd ACM Workshop Security of Ad Hoc and Sensor Networks, pp. 89–96. ACM Press (2007)
3. Wood, A.D., Stankovic, J.A.: Denial of service in sensor networks. Computer **35**(10), 54–62 (2002)
4. Raymond, D.R., Midkiff, S.F.: Denial-of-service in wireless sensor networks: attacks and defenses. IEEE Pervasive Comput. **7**(1), 74–81 (2008)
5. Xu, W. et al.: The feasibility of launching and detecting jamming attacks in wireless networks. In: Proceedings of 11th Annual International Conference on Mobile Computing and Networking, pp. 46–57. ACM Press (2005)
6. Woo, A., Tong T., Culler, D.: Taming the underlying challenges for reliable multi-hop routing in sensor networks. In: Proceedings of 3rd ACM International Conference on Embedded Networked Sensor Systems, pp. 14–27. ACM Press (2010)
7. Raymond, D. et al.: Effects of denial of sleep attacks on wireless sensor network MAC protocols. In: Proceedings of 7th Annual IEEE Systems, Man, and Cybernetics (SMC) Information Assurance Workshop (IAW), pp. 297–304. IEEE Press (2006)
8. Stajano, F., Anderson, R.: The resurrecting duckling: security issues for ad-hoc wireless networks. In: Christianson, B., Crispo, B., Malcolm, J.A., Roe, M. (eds.) Security Protocols 2000. LNCS, vol. 1796, pp. 172–182. Springer, Heidelberg (2011)
9. Crosby, G.V., Pissinou, N., Gadze, J.: A framework for trust-based cluster head election in wireless sensor networks. In: Proceedings of 2nd IEEE Workshop Dependability and Security in Sensor Networks and Systems, pp. 13–22. IEEE Press (2006)
10. Jakobsson, M.: Fractal hash sequence representation and traversal. In: 2002 IEEE International Symposium on Information Theory (ISIT 2002), Switzerland, July 2002
11. Liu, D., Ning, P.: Efficient distribution of key chain commitments for broadcast authentication in distributed Sensor Networks. In: NDSS (2003)
12. Zhu, S., Setia, S., Jajodia, S., Ning, P.: An interleaved hop-by-hop authentication scheme for filtering of injected false data in sensor networks. In: 2004 IEEE Symposium on Security and Privacy, Oakland, CA, USA, May 2004

Swarm Intelligence Based Data Aggregation for Intruder Detection in Wireless Sensor Networks

Muthuraman Thangaraj[1(✉)] and Pichaiah Punitha Ponmalar[2]

[1] Department of Computer Science,
Madurai Kamaraj University, Madurai, India
thangarajmku@yahoo.com
[2] Department of Computer Science,
Sri Meenakshi Govt Arts College, Madurai, India
p_punithanadraj74@yahoo.com

Abstract. Wireless Sensor Network in which nodes are mobile can be defined as MWSN (Mobile Wireless Sensor Networks). MWSNs are a budding field of research in contrast to their well-established ancestor. It consists of a variety of sensing elements deployed indiscriminately and at enormous scale. This can build a huge variety of redundant sensory information. Transmission of such redundant data saturates network resources and it additionally consumes nodes energy. Data aggregation is the process of reducing the amount of packets to be sent to sink to increase the life time of MWSN. Path and location of the nodes are changing frequently in mobile WSN. So Swarm intelligence (group behaviour) is needed for the communication as well as on data collection, Query publication and intrusion detection. To operate MWSNs in a secure way, intrusions should be detected before attackers can harm the network. Intrusion can also occur in multiple layers of MWSN so continuous investigation of aggregated data to be needed. Proposed work is for SIDI (Swarm Intelligence based Data aggregation for Intruder detection) raising the energy efficiency of a network. Simulation results indicate that this operation practically will increase the network lifetime.

Keywords: Wireless sensor network · Data aggregation · Swarm intelligence · Intrusion detection

1 Introduction

Wireless Sensor Network (WSN) consists of a set of sensor nodes launched in a distributed fashion which are connected to the node called sink. Mobile Wireless Sensor Networks [10] (MWSNs) have recently launched a budding popular class of WSN in which mobility plays a key role in the implementation of the application. The use of the network is to monitor the environmental factors like humidity, temperature and pressure. The application area of WSN is in the industry, weather forecasting, wild fire detection, smart homes, traffic control, military and in agriculture. In several cases the nodes are thrown to the remote area and it is difficult to charge the battery of the

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 588–601, 2015.
DOI: 10.1007/978-3-319-21009-4_44

nodes. The lifetime of the sensor depends on its energy. Data propagation cycle of the sensor consumes more energy. Minimizing the propagation of data will improve the energy loss. Swarm Intelligence techniques are applied, thereby improving energy efficiency and effectiveness. It reduces the massive amount of distributed log data moved among the inner nodes. Having agents visited the data repositories and mine results. In Mobile Wireless sensor Networks Alien data is dangerous. Before collecting data from the nodes, agents check for intrusion. Then it collects the data from multiple sensors. The data from multiple sensors produces redundancy. To eliminate redundancy, data aggregation is used to aggregate the data from multiple sensors.

The rest of the paper is organized as follows. In Sect. 2 Background and Related works are discussed and proposed work is in Sects. 3 and 4 discusses simulation results and finally in Sect. 5 the paper is concluded.

2 Background and Related Work

In Wireless sensor network (WSNs) nodes are massively deployed in a region of interest to collect information from their surroundings In recent years, a large amount of WSN-related applications such as object tracking, health monitoring, security surveillance, and intelligent transportation have been proposed. Usually WSN [8] is deployed with static sensor nodes to perform monitoring missions. However, due to the dynamic changes of events and hostile environment, a pure static WSN could face the following challenges like connectivity, coverage and energy consumption. In recent years, mobility has become a major area of research for the WSN commune. MWSN [10] have recently launched a budding popular class of WSN in which mobility plays a key role in the implementation of the application. MWSN allows the sensor nodes to move freely and they are able to communicate with each other without the need for a fixed infrastructure. Mobile networks are capable of out-performing static wireless sensor networks as they tend to increase the network lifetime, reduce the power consumption, provide more channel capacity and perform better targeting in Fig. 1.

The increasing capabilities and the decreasing costs of mobile sensors make mobile sensor networks possible and practical. Data query and publication and path definition are volatile in mobile wireless sensor networks.

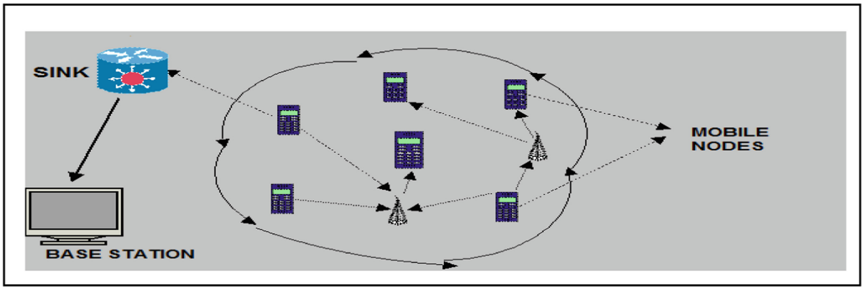


Fig. 1. Mobile wireless sensor network

Mobile Sensor nodes in the corresponding spatial portion of the network act as event-based publishers. They publish the information towards the subscribing sink if and when they detect the indicated phenomenon. However since various nodes often detect common phenomenon, there is likely to be some redundancy in the data. In-network filtering and processing techniques can help to conserve the scarce energy resources.

Data aggregation [9] has been put forward as an essential paradigm for MWSN. The idea is to aggregate the data coming from different sources to eliminate redundancy, reduce the number of transmissions and thus save energy. Data aggregation is varying from data pre-processing. Data aggregation is the process of aggregation of information to produce combined data. Data pre-processing is the process of cleaning, normalization, transformation, feature extraction and selection etc.

In Mobile Wireless sensor Networks Alien data is dangerous. There aren't any gateways or switches to watch the data flow, the protection of such networks could be a big concern. It needs continuous investigations of aggregated data in multiple layers. Intrusion Detection is the detection of any suspicious behavior in a network. To collect data from the nodes and to find the intrusion in the nodes needs mobile devices or mobile agents.

A swarm can be viewed as a group of agents cooperating to achieve some focused behavior and achieve some goal. Swarm Intelligence [2] is a nature inspired computational model based on the collective behavior. There are numerous applications of swarm intelligence; one among them is routing in communication networks. Routing based on swarm intelligence provides a promising alternative to the traditional routing approaches. Since it uses mobile software agents for network management, which have the ability to adapt, cooperate and move intelligently from one location to the other. It reduces the massive amount of data moved among the inner nodes. Having agents visited the data repositories and mine results is an ideal alternative. These agents are used in MWSN for communication, query publishing and intrusion detection.

2.1 Related Work

Amundson [1] proposed Mobile sensor localization and Navigation using Doppler shifts enforced on a MWSN network. Networks of mobile sensors require position updates for navigation through the sensing region.

Dervis Karaboga, and Beyza Gorkemli [2] discussed about Probabilistic Dynamic Deployment of Wireless Sensor Networks by Artificial Bee Colony Algorithm. They discussed about the artificial bee colony algorithm is applied to the dynamic deployment of stationary and mobile sensor networks. Dervis Karaboga and Bahriye Basturk [3] proposed a powerful and efficient algorithm for numerical function optimization using Artificial Bee Colony (ABC) algorithm. Cluster based wireless sensor network routing using artificial bee colony algorithm proposed by Dervis Karaboga et al. [4]. Clustering is one of the most popular techniques preferred in routing operations. In this work, a novel energy efficient clustering mechanism, based on artificial bee colony algorithm, is presented to prolong the network life-time.

Vladimir Dyo and Cecilia Mascolo [5] proposed the Efficient Node Discovery in Mobile Wireless Sensor Networks. A large portion of the energy of sensor applications is spent in node discovery as nodes need to periodically advertise their presence and be awake to discover other nodes for data exchange.

Erfu Yang and Ahmet [6] proposed an Improved Particle Swarm Optimization Algorithm for Power-Efficient Wireless Sensor Networks. This paper presents an improved Particle Swarm Optimization (PSO) algorithm for onboard embedded applications in power-efficient wireless sensor networks. Muhammad Saleem and Muddassar Farooq [7] discussed about Swarm intelligence based routing protocol for wireless sensor networks.

Javad Rezazadeh et al. [10] discussed Mobile Wireless Sensor Networks Overview. In this article, an overview of proposals that evaluate mobile communication in WSNs is presented.

R. Sagayam [11] compared Ant Colony and Bee Colony optimization for spam host detection and they presented a spam host detection approach.

Xiangyu Yu et al. [12] proposed a faster convergence Artificial Bee Colony Algorithm in Sensor Deployment for Wireless Sensor Networks.

3 Proposed Scheme

Mobile sensor networks can be classified into three layer network architectures in Table 1. MWSN need mobile devices to collect data from the node. Single mobile devices or single mobile agent is not enough to satisfy the need. Multiple mobile agents are needed. If multiple agents collect the data from multiple nodes in the same region will cause redundancy and the agent cannot differentiate between malicious node and the actual node.

Table 1. MWSN Layer Architecture

Layers	Architecture	Devices	Communication
layer 1	Flat or Planar	Mobile Nodes or Stationary Nodes	Same network
layer 2	Overlay	Set of Stationary or set of Mobile nodes(data mules)	all nodes are always connected
layer 3	Heterogeneous	Set of stationary nodes and mobile devices	Stationary to mobile then mobile to access point

Proposed work Swarm intelligent based data aggregation for intruder detection for MWSN (SIDI) deals these issues. Swarm intelligence is for creating multiple agents and to collect data from mobile nodes. Data aggregation is used to reduce the redundancy in the collected data and intruder detection to identify malicious node or data. Swarm intelligence based agents are used to collect data from sensor nodes in Fig. 2. It delivers dedicated support for bio-inspired features such as transparent agent migration and pheromone trails. A widely employed bio-inspired paradigm that has been applied to computer networks is the mimicking social insects such as ants and bees.

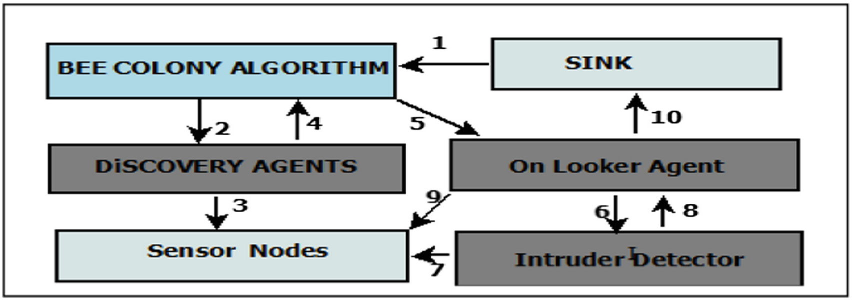


Fig. 2. SIDI components

In our proposed work Bee colony algorithm is applied to collect the information from the nodes. Then network represents the environment where insect like software agents, each node represents nest that accepts incoming agents and agents themselves are represented by messages exchanged between nodes. Behavior of nodes and exchanged information is defined by means of algorithm. Agents assigned with different tasks such as discovering distant nodes, connecting and disconnecting peers, or signaling the aliveness status of node.

The proposed SIDI architecture in Fig. 3 is for Swarm Intelligence Based Data aggregation for Intruder detection (SIDI) is to improve the energy efficiency, channel capacity and data fidelity. It includes three modules such as swarm intelligence module, intrusion module and aggregation module.

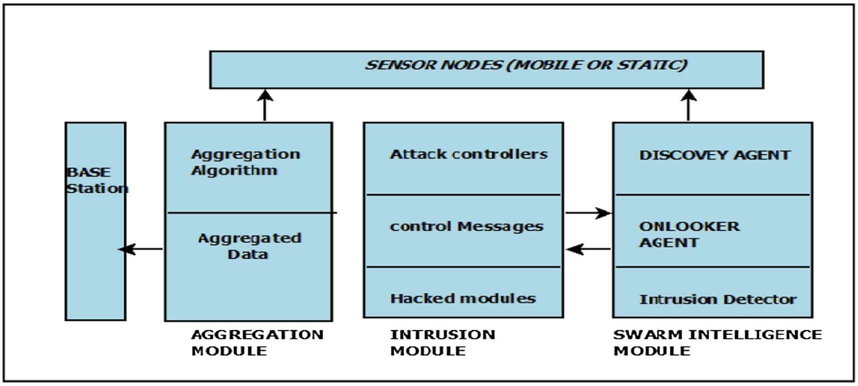


Fig. 3. SIDI architecture

Figure 4 explains the workflow of the Architecture. Swarm intelligence module include the agents such as

1. Discovery Agent
2. Onlooker Agent
3. Intruder Detection

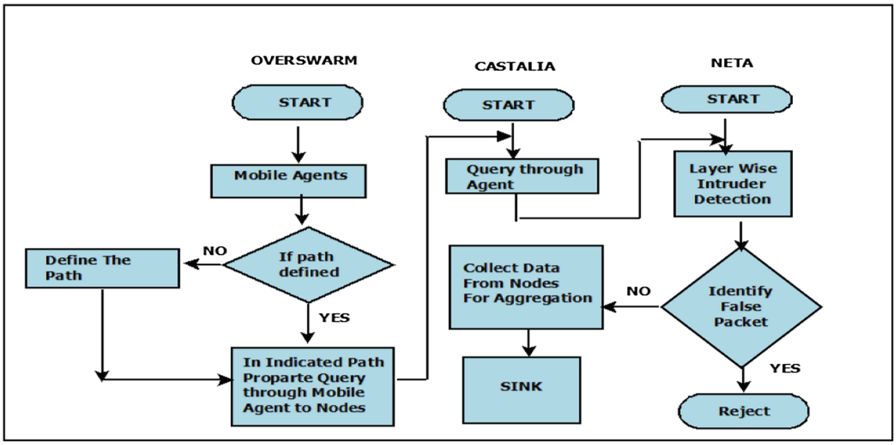


Fig. 4. Workflow of SIDI

Discovery agent determines distant mobile nodes placed on the network. Based on the information collected by discovery agents, Onlooker agent propagates through that path to collect the data from Nodes. Onlooker agent data's are checked for intruder's attack.

Intrusion module includes Attack controllers, Control messages and hacked modules. Using of Attack controllers intrusions are identified if any intruder or false packet is identified in any layer this module informs it to the swarm intelligence module. After removing false data, onlooker agent collects and aggregates the data and transmits it to the sink.

3.1 Discovery Agent

In Table 2 Discovery agents are periodically instanced on each node with some probability. Each agent wanders on the overlay for a predefined maximum number of steps. The list of sources (Source List) specifies what the sources to be visited by the DA and the list is stored in bounded size vector. There are two special sources, namely, the first source (First Source) and the last source (Last Source). Working principles of Discovery agent is explained in algorithm.

Table 2. Discovery agent packet structure

SINK ID	DA-SEQ NUM		OPTIMIZATION LINK	
	FIRST SOURCE	NEXTHOP	LAST SOURCE	SOURCELIST
PROCESSING CODE	ENERGY VALUES	CONSTRUCT	ESTIMATED DISTANCE	

The pair of First Source and Last Source indicates the beginning and end points of the DA data gathering. Next Hop indicates the immediate next hop node which is an intermediate sensor node or a target source node. Next hop is chosen among the neighbors of the current peer. The agent migrates to the neighbor associated with the lowest concentration of gamma pheromone; otherwise it migrates to a random neighbor. Optimization Link is to create logical link between nodes in order to optimize the topology of the overlay to its target.

Estimate Distance returns the distance from the current node to the source. Construction is used to connect new peers to the overlay as well as to recover connectivity in the event of a failure. DA-SEQ NUM to specify the sequence number for the agents and the source list is to record the visited nodes. Energy value field is to store the remaining energy value of the nodes.

Algorithm Discovery Agent (Input: Source List, Output: Node info, Energy)

*Check whether an DA arrives at a specific source or not */*
(Node=First Source)
DA migrates toward First Source;
DA calculates Estimated Distance;
Repeat
    Record the sources in Source List, select the Next Hop;
    Set Next Hop as Next Source;
    DA calculates Estimated Distance
    DA migrates towards Next Source.
***Until** (Node=! Last Source)*

3.2 Onlooker Agent

The itinerary of the agent is already set up before the sink dispatches it. Path information placed into this agent. The onlooker agent is to carry processing codes that allow the computation and communication resources at the sensor nodes. The processing code carried by the OA packet only requires local processing of the raw data at the source nodes. This capability enables a reduction in the amount of data transmission by allowing only relevant information to be extracted and transmitted. Let r, $(0 < r < 1)$, be the data reduction ratio contributed by the OA assisted local processing, S be the size of Onlooker agent Packet Structure The information contained in an OA packet is shown in Table 3. The pair of Sink ID and OA_SeqNum is used to identify an OA packet. Whenever a sink dispatches a new OA packet, it will increment the OA_SeqNum. The list of sources (Source List) specifies which sources to be visited by

Table 3. Onlooker Agent Packet Structure

SINK ID	OA-SEQ NUM		OPTIMIZATION LINK	
	FIRST SOURCE	NEXTHOP	LAST SOURCE	SOURCELIST
PROCESSING CODE	DATA			

the OA. The first source (First Source) and the last source (Last Source) indicate the beginning and ending visit of the OA. Next Hop indicates the immediate next hop node which is an intermediate sensor node or a target source node. The agent migrates to the nodes and the values are updated in Source list. The payload of an OA packet includes Processing Code. It carries the query for the data. Data field is for the aggregated data.

Working principle of onlooker agent is explained in algorithm. Onlooker agent algorithm explains how an OA migrates in a round. When the OA arrives at a sensor node, it looks at the identifier of the current node to decide whether or not it has arrived at the destination source. If not, the OA continues migrating towards the specific source.

Algorithm: Onlooker Agent (Input: Source list, Output: Aggregated Data)

Node=First Source
OA migrates toward First Source;
OA collects data;
TAG= IDS(data)
If TAG =1
OA aggregates the data
Repeat
    Among the sources in Source List, select the Next Source;
    Set Next Source in the OA packet;
    OA migrates towards Next Source.
    OA collects data;
    TAG= IDS (data)
    If TAG=1
    OA aggregates the data
***Until** (Node=! Last Source or (All nodes are visited from the Source list))*

3.3 Intrusion Detector

IDS main components are: 1. *Attack Controllers:* These controllers are used to manage attack node by sending control messages. The fields of controls messages are Attack type, Active, Start time, End Time, Attack specific parameters.

2. *Control Messages:* These messages are sent from attack controllers to the hacked modules involved in the attack execution. They transmit the information necessary for activation and deactivation of the attacks.

3. *Hacked Modules:* These are the modules whose behavior is modified in order to strike an attack. Malicious nodes will inject alien data. This will affect QOS parameters.

Algorithm: IDS (Input: Source list, Last source data, Output: DATA, TAG)

Repeat
  Send control message to detect layer wise intrusion
  if intrusion occurs reject the packet and TAG=0
    Else
    Collect data from the node and TAG=1
***Until** (Node! =LAST SOURCE)*

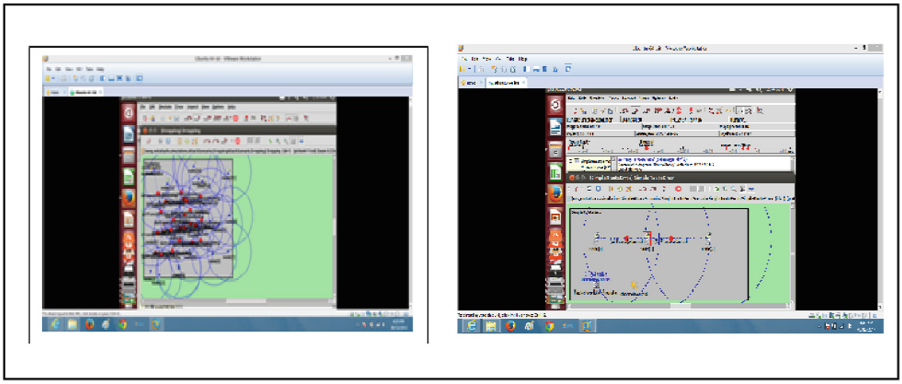


Fig. 5. NETA framework Screen shot

4 Simulation Results

The simulations of SIDI are performed via OMNET simulator using INET and Mixim framework, NETA, OVERSWARM and CASTALIA. By varying the number of SNs i.e. 5, 10, 15, SNs which are randomly distributed over a 600 × 400 m. Mobility Model is Random way point, Radio Sensitivity is −85 mW and Bit rate is 2 Mbps. Figure 5 shows the NETA framework screen shot and Fig. 6 shows the CASTALIA and OVERSWARM framework screen shot.

The formula for producing a candidate solution from the existing is described below

$$k_{ij} = \gamma_{ij}\left(y_{ij} - y_{kj}\right) \tag{1}$$

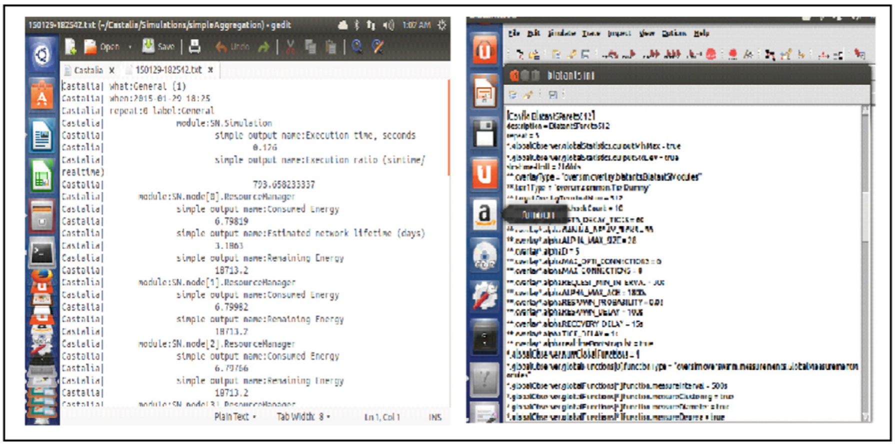


Fig. 6. CASTALIA and OVERSWARM framework Screen shot

$$k_{ij} = \gamma_{ij}(1 + y_{ij}) - \gamma_{ij} y_{kj} \tag{2}$$

$$m_{ij} = y_{ij} + k_{ij} \tag{3}$$

Where k: {1, 2,….. Number of Employed Bees}
j: {1, 2,….L} is randomly chosen index
L: Number of parameters to optimize
k <> i: both are randomly chosen,
$\Upsilon_{i,j}$: random number [− 1, + 1]

After each candidate source position mij is produced and then evaluated by the artificial bee, its performance is compared with that of Yij. If the new food has equal or better nectar than the old source, it is replaced with the old one in the memory. Otherwise, the old one is retained.

Bee.Probabilities: An onlooker bee agent chooses a food source depending on the probability value associated with that food source, prob (i), calculated by:

$$fitness(n) = \sum\nolimits_{l=1}^{n} fitness(l) \tag{4}$$

$$prob(i) = \frac{fitness(i)}{fitness(n)} \tag{5}$$

The Performance outcome of the genetic algorithm applied in this work would affect an efficient presentation in minimizing the energy consumption.

5 Performance Analysis

Table 4 Shows that the Battery residual capacity of data aggregation using GA and SIDI. With the increase in the number of iterations, Energy utilization of SIDI decreased and energy utilization of GA increased. Lifetime of the sensor node depends on the remaining energy of the sensor node. After applying SIDI algorithm the residual capacity the node remains high.

Table 4. Battery residual capacity

No. of iter ations	Battery capacity (amp)			Percentage of energy utilized	
	Default	Trad GA	SIDI	GA	SIDI
1	***10***	***8***	***9***	***20***	***10***
2	***9***	***5***	***7***	***40***	***20***
3	***8***	***2***	***7***	***50***	***10***

Table 5 shows comparison of the detection rate of SIDI and GA. Table 6 shows the effectiveness of SIDI by comparing with other data aggregation algorithms. Comparison includes the parameter such as detection rate; false positives and false negatives.

5.1 Detection Rate and Performance Measurement

5.2 End to End Delay

After applying algorithm SIDI the End-end delay is reduced. The time interval between the information packets is received by SN until the time instant when the packet is delivered to the BS. The average delay of SIDI and energy efficient GA is highlighted in Fig. 7.

Table 5. Comparison of detection rate

NODES	DETECTION RATE PERCENTAGE			
	TP	TN	GA	SIDI
2	1	1	65	50
3	2	1	70	66
5	3	2	72	60
10	7	2	80	77
15	7	3	78	70

Detection Rate is calculated as follows
DR = TP/(TP + TN) X 100 %
TP = amount of attack when it actually attack
TN = amount of normal detect when it actually normal

Table 6. Comparison of performance measurement

PARAMETERS	LEACH	PEGASIS	SPIN	SIDI
Detection Rate	High	Medium	High	Low
False Positive	High	High	High	Low
False Negative	High	High	High	Low
End-End Delay	High	High	High	Low
Computational Cost	Medium	Medium	High	Low
Energy Consumption	High	High	High	Low
Fault Tolerance	High	High	Medium	Low

5.3 Residual Energy

Figure 8 Shows that Residual energy ratio between GA based data aggregation and SIDI. Residual Energy means the remaining energy after the packets are delivered from node to sink. With the increase in the number of iterations, Residual energy is more for SIDI than GA.

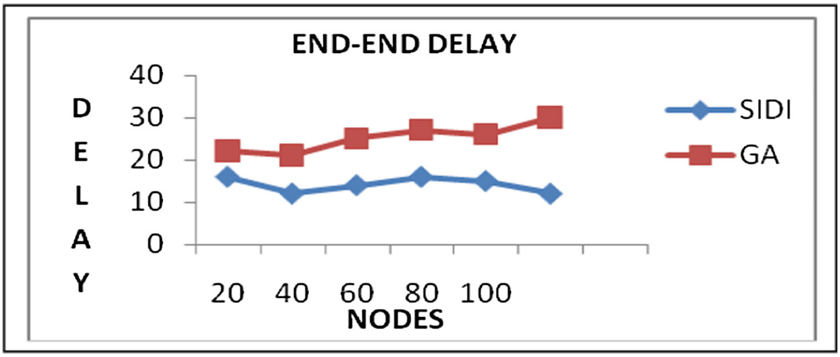


Fig. 7. Nodes Vs Delay

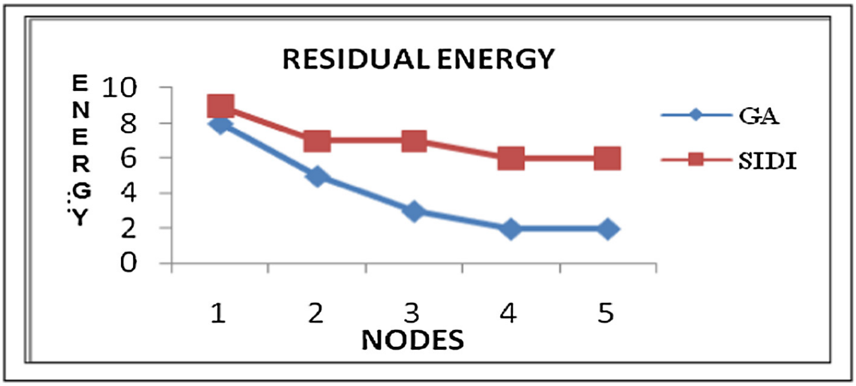


Fig. 8. Nodes Vs Residual energy

5.4 False Negative and False Positive

Figure 9 shows that False Negative and False Positive ratio between Data aggregation algorithm SIDI, GA and LEACH. A false Negative is an error in data reporting in which a test result improperly indicates no presence condition when it is really present. A false positive is an error in data reporting in which a test result improperly indicates presence condition when it is really not present.

5.5 Findings into Practice

SIDI algorithm can be used in healthcare application, tracking moving objects, Engineering and industries. In health care application patient details are collected from the wearable or non wearable sensors through agents of SIDI. Before aggregation the collected data are checked for intrusion. IDS is not included means intruder can alter the medical data it will create a serious problem to the patient.

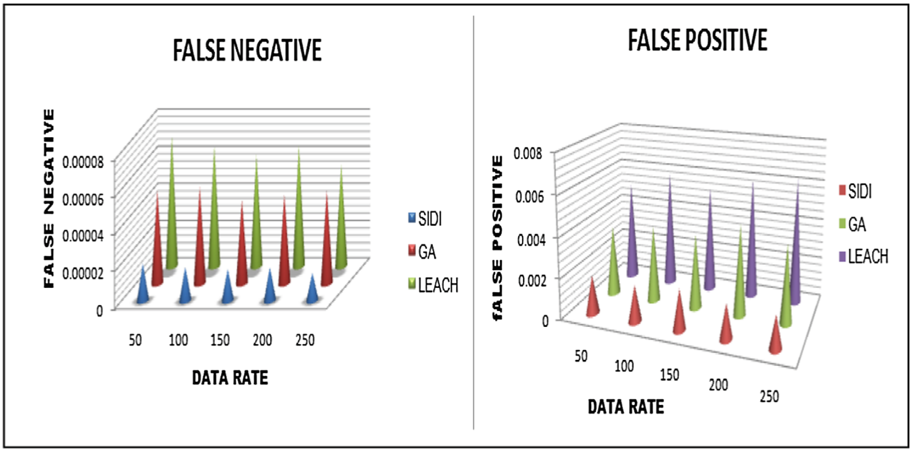


Fig. 9. False negative and false positive ratio

6 Conclusion

Wireless sensor network consist of large sets of nodes. The design of effective routing protocols in these networks is a challenging task. Swarm intelligence offers algorithmic design principles, taking inspiration from foraging behaviours. In most of the swarm-based aggregation techniques, only the energy consumption is considered for data aggregation. It is easy for an opponent to inject false data into the network and misinform the aggregator to accept false readings. So, the need for secure data aggregation is necessary in MWSN. An innovative SIDI algorithm was introduced which improves the lifetime of WSN effectiveness and increases the system's accuracy. Swarm intelligence was used to choose the efficient path, locate the intruders and Bee colony agents to aggregate the data. It provides reliable data gathering, Scalability, extensive energy savings, accuracy and security. SIDI can be applied in Health care system, Moving object tracking, Engineering and Industries. Simulation results show that this method is more efficient in extending the lifetime of MWSN than Genetic algorithm based method. Further investigations may include the use of other intelligent algorithms instead of Bee colony.

Acknowledgements. We would like to thank anonymous reviewers whose careful reading and constructive criticism of earlier draft help to improve the clarity and the content of this paper.

References

1. Amundson, I., Koutsoukos, X., Sallai, J., Ozturk, C.: Mobile sensor localization and Navigation using RF doppler shifts. In: Ist ACM International work shop on mobile entity localization and Tracking in GPS-less Environments, MELT (2008)

2. Karaboga, D., Gorkemli, B.: Probabilistic dynamic deployment of wireless sensor networks by artificial bee colony algorithm. Sensors **11**(6), 6056–6065 (2011)
3. Karaboga, D., Basturk, B.: A powerful and efficient algorithm for numerical function optimization: Artificial Bee Colony algorithm. Glob. Optim. **39**(3), 459–471 (2007)
4. Karaboga, D., Okdem, S., Oztur, C.: Cluster based wireless sensor network routing using Artificial Bee Colony algorithm. Wireless Netw. **18**, 847–860 (2012)
5. Dyo, M.: Efficient node discovery in mobile wireless sensor networks. In: Proceedings of 4th International Conference on Distributive Computing in Sensor Systems, pp. 157–475 (2008)
6. Yang, E., Erdogan, A.T., Arslan, T., Barton, N.: An Improved Particle Swarm Optimization Algorithm for Power-Efficient Wireless Sensor Networks (2007). IEEE 0-7695-2919-4/07
7. Di Caro, G.A., Farooq, M.: Swarm intelligence based routing protocol for wireless sensor networks: Survey and future directions. Inf. Sci. **181**, 4597–4624 (2011)
8. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor networks: a survey. Comput. Netw. **38**, 393–422 (2002)
9. Thangaraj, M., Punitha Ponmalar, P.: A survey on data aggregation techniques in wireless sensor networks. Int. J. Res. Rev. Wireless Sensor Netw. (IJRRWSN) **1**(2), 36–42 (2011)
10. Rezazadeh, J., Moradi, M., Ismail, A.S.: Mobile wireless sensor networks overview. IJCCN Int. J. Comput. Commun. Netw. **2**(1), 17–22 (2012)
11. Sagayam, R.: Comparison of Ant Colony and Bee colony for spam host detection. Int. J. Eng. Res. Develop. November 2012. ISSN 2278-06/x, p:ISSN 2278-800x
12. Yu, X., Zhang, I., Fan, I., Zhan, T.: A faster convergence Artificial Bee Colony algorithm in sensor deployment for wireless sensor networks. Int. J. Distrib. Sensor Netw. **2013**, 9 (2013)

Intelligent Sybil Attack Detection on Abnormal Connectivity Behavior in Mobile Social Networks

Anand Chinchore(✉), Frank Jiang, and Guandong Xu

Advanced Analytics Institute, University of Technology Sydney, Ultimo, Australia
anand.a.chinchore@student.uts.edu.au,
{frank.jiang,guandong.xu}@uts.edu.au

Abstract. There have been a large number of researches on mobile networks in the literature, focusing on a variety of secured applications over the network, including the use of their connections, fake identification and attacks on social group. These applications are created for the intention to collect confidential information, money laundering, blackmailing and to perform other crime activity. The purpose of this research is to identify the behavior of the honest node (network account) and fake node (network account) on mobile social network.

In this research, the behavior survey of these nodes is carried out and further analysed with the help of graph-based Sybil detection system. This paper particularly studies Sybil attacks and its defense system for IoT (Internet-of-Things) environment. To be implied, the identification of each forged Sybil node is to be tracked on the basis of nodes connectivity and their timing of connectivity as well as frequency among each other. Sybil node has a forged identity in different locations and also reports its virtual location information to servers.

Keywords: Sybil attack · Mobile social network · Anomaly detection

1 Introduction

Over the last decade, a large amount of complicated and different structured data has been generated and flooded on the internet, and it has the increasing trend in sizes and dimensions nowadays. Users of mobile technology and Apps are quickly occupying the personalised services market. Uploading data, usage of social networking, banking systems, storage and production of text data, images, animations and storage of confidential information on mobile become as much easier than that was on computers and laptops. Spreading and transferring such data through the mobile and social network are also easier than that via emails. It is a time-saving and quickest process as human psychology and behavior have been recognized.

Additionally, the compression of real data also becomes easier due to Smartphone technology. It is free and quick rather using computer software. Free Apps and inbuilt mobile technology made the human social, financial, cultural and official life easier than before.

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 602–617, 2015.
DOI: 10.1007/978-3-319-21009-4_45

Meanwhile, this technology makes the attacker work much simpler and facilitates the attacker to steal such information from mobile social network, groups in fraction of seconds without using any special hacking software. Such attackers are called a Sybil attacker on mobile social networks, who pretends to be an honest entity and collect such information from real legitimate entity's intention to for blackmailing, money laundering, hacking, and to perform other criminal and offensive activities. This data may take for self-use or distribute among other Sybil attackers.

The honest node is the node which is real and connected to other honest nodes. The false node is one with the fake identity or profiles those connected to other honest nodes to steal information. Fake nodes are also connected to other fake nodes with the intent to exchange honest node data collected between other fake nodes. This behavior will be tracked with the help of connectivity between an honest node and a fake node and it's time of connection with each other.

Sybil users are those who create a fake identity on mobile social network for the intention of data distribution. A single user or group of Sybil's may harm the social media or particular organization or an individual. The intention behind activity may intend to reduce the revenue of the organisation, stealing the customer data and sell it to other group, taking down the social media provider to decrease the user involvement new user and their registration, or may be misuse the data of individual to achieve personal aim.

The honest user accesses its identity on mobile social network for the purpose of personal use of communication. The honest entity may connect one or more trustworthy entity including few unknown Sybil identities. On the other hand, one Sybil user may have more than one identity on mobile social network. These identities can or may access from one mobile device or multiple mobile devices. The purpose of this is again to interfere in other personal honest entities profiles for information uploaded by honest identity.

In addition, two or more Sybil users may get to gather and create a group or community of Sybil's. The purpose to form this community is to generate a attack to harm individual or organizations at a very high level of destruction of data or information. This group of people could be dangerous in performing a Sybil attack on an honest community. This attack could be performing by pretending as an honest user in an honest community by Sybil user(s). One or more Sybil users may get involve into the honest community pretending as an honest user and pass the information to the other Sybil community user into the Sybil community or honest community. These kinds of profiles or identities are hard to guess amongst the honest region as they were just the same as honest identities.

Furthermore, the Sybil user generates connection with other Sybil users(s) from honest region to Sybil region. This connection is sometimes hard to identify as Sybil entities could be using a changing IP location or fake generated IP addresses by special tools and software (Zhang et al., 2013).

To identify such Sybil user for their abnormal behavior, not many research work have been published. The intention of performing such activity is common among the people as per the human psychology. This kind of behavior may intend personal damage of assets or because of the past harmful mind-disturbing events in real life of Sybil or it may be just for fun. These events could be judged by the individual to individuals or by the other people in the same community. Sybil user performs same actions and things in a virtual world.

The abnormal behavior could be observed as the following scenarios: breaking up between couples, resisting attack, money laundering, stealing individual information for the personal use. as well as performing terrorist attack in regions, collecting data from multiple users to sell to other organizations to generate revenue, breaking down the revenue of the organization, harm to individual life or organization from personal point of view or based on incident in a past life, or it could be because many human psychological barriers. Such behavior ranges from low intensity to very high intensity, which is just stealing data from individuals up to the terrorist attack in an actual world region through social media. Some of the behavior could be seen as childish acted by some Sybil user because of basic level of educational information to just to make fun of other honest identities. These users mostly categorized under the age 14–22 years old and perform actions due to the lack of knowledge about Law.

The research aims at identifying Sybil user or a group of Sybil user in a complicated environment. The purpose of this study is to evaluate a data set and find the connection among the users to identify the suspicious activity performs by fake identities to improve social media and reduce its threats for individual or organization. As there are multiple information sources exploited to enrich the quality and quantity of researches by same domain, this study area plays a significant role in Social Media Analytics. To summarize, the research proposed in this paper resolves the following issues:

(1) Does the abnormal behavior between nodes on mobile social network, identify from the connection that they have made with honest node and Sybil node?
(2) To identify this abnormal behavior, does the irregular time of connection between nodes also plays a vital role?
(3) Does the connectivity intensity or the number of times of connections can help to identify Sybil?

The above research questions have been taken into consideration of the dataset availability and the records in it. The dataset shows User 1 connection (total 99 users) and connection to user 2 (over 4000 users) connected multiple times to user one. Each connection has a starting time, ending time and its index value.

The above research problems will be evaluated by the dataset from Infocom 2006, and will produce the results based on objectives specified in this research. The final outcome will be to explain in the form of algorithm, graphs, charts and the table which will display each node containing assumption, Sybil category, reason, and each node connected other nodes.

The research aims at using the developed graph-based system to determine the connectivity intensity or the number of times of connections of honest nodes to Sybil node and Sybil nodes to other Sybil nodes.

Aim 1 Is to identify Sybil user or a group of Sybil users in an honest region
Aim 2 The aim of this research is to evaluate the data set and find a connection among the users to identify the suspicious activity that performs by fake identities

Aim 3 Improve social media and reduce its threats for an individual or organization (s). Also, it will be an addition to Social Media Analytics researchers

This research thesis explains Sybil attacks can be traced with the help of graph-based system and behavior analytics system. To identify the Sybil user in an honest region by cross checking method that could be a number of times connection performed with a single user or multiple users. To show and explain this practically it will identify the suspicious activity acts by fake identities, one, regular or continuous connection with single or multiple users for longer time which will helps to identify Sybil user and second, the irregular time of connection between two nodes creates suspicious activity between nodes. For programming this algorithm to get into use and improve social media and reduce its threats for individual or organization(s), developed graph-based system helps to identify the connectivity intensity or the number of times of connections between honest node to Sybil node and Sybil nodes to other Sybil nodes will be use MATLAB and SPSS in this research.

Significance 1 The research identifies the fake identity using graph-based technique

Using graph-based techniques in this research the user1, user 2, starting time and ending time of the dataset infocom06 will produce the different graphs and charts to identify Sybil users. It will also evaluate the relationship between the nodes based on the number of connection takes place with each user. Then further, this graphs help to identify the Sybil attacker and behavior of the Sybil node.

Significance 2 To track the behavior of Sybil node based on time of connectivity between each user and number of times connection with a single user

Using the behavior technique, the research outcome will enable to identify the behavior of each user by starting time and ending time of connections. This will help to improve the results of significance 1 and will achieve maximum success to trace the Sybil user in honest community.

This paper is organized as follows: Sect. 2 describes the existing research and related literature review on Sybil attacks and their behavior issues. Section 3 explains the research method and research objectives produced using designed algorithms which shows output in the format of bar graphs and tables. This methodology explains the connection based behaviors between nodes. Section 4 highlights the main discussion points with conclusion followed by references.

2 Related Works

The evolution of sybil defense protocols that leverage the structural properties of the social graph underlying a distributed system to identify sybil identities. Alvisi L. and team first, clarify the deep connection between sybil defense and the theory of random walks which leads to identify a community detection algorithm that, for the first time, offers provable guarantees in the context of sybil defense. Second, team advocate a new goal for sybil defense that addresses the more limited, but practically useful, goal of securely white-listing a local region of the graph (Alvisi et al., 2013).

In one of the research SyMon: A practical approach to defend large structured P2P systems against Sybil Attack researcher Jyothi B. S. and Dharanipragada Janakiram explains the chosen SyMon entrusted with the responsibility of moderating the transactions involving the given peer and hence makes it almost impossible for sybils to compromise the system. Research shows the effectiveness of the proposed system in defending against Sybil attack both analytically and experimentally in large structured P2P System.

A case study in an Australian mobile network operator is presented in order to highlight the decision-making process for the launch of mobile networking services. By critical analyzing the insights gained from the competitive, customer, and value delivery analysis, it was possible to derive a three-step strategy that MNOs can adopt regarding the offer of mobile social networking services: (1) extend SN access to mobile; (2) enrich and differentiate SN experiences by MNO service integration; and (3) drive usage & acquisition of MNO services with SN integration (Cortimiglia et al., 2011).

SybilGuard protocol leverages a key insight on social networks to bound the number of sybil nodes accepted. Research paper presents the novel SybilLimit protocol that leverages the same insight as SybilGuard, but offers dramatically improved and near-optimal guarantees (Haifeng Yu et al. 2010).

Research Paper presents SybilGuard, a novel protocol for limiting the corruptive influences of sybil attacks. Protocol is based on the “social network” among user identities, where an edge between two identities indicates a human-established trust relationship. Thus, there is a disproportionately small “cut” in the graph between the sybil nodes and the honest nodes. SybilGuard exploits this property to bound the number of identities a malicious user can create(Haifeng Yu et al. 2008).

Temporary connection cannot be recognized as a real relationship when the history connection among the nodes as considered. Cumulative stable contacts are proposed to depict the correlation among the nodes. With each timestamp, the community core detected due to the variation of nodes and links. The propose research method effectively detect stable community in mobile social network (Hao Xu et al., 2013).

Evolution of person-to-person social relationships, quantify and predict social tie strengths based on call-detail records of mobile phones. Researcher propose an affinity model for quantifying social-tie strengths in which a reciprocity index is integrated to measure the level of reciprocity between users and their communication partners (Zhang et al., 2010).

Huiqi Zhang and team propose a Socioscope model for social-network and human-behavior analysis based on mobile phone call-detail records. Researcher used multiple probability and statistical methods for quantifying social groups, relationships, and communication patterns and for detecting human-behavior changes including new index to measure the level of reciprocity between users and their communication partners (Zhang et al., 2011).

Jiang J. and team (2012) produced a research on detecting and validated the Sybil groups in online social network. As per their consideration, Sybil users propagate spam or unfairly increase the influence of target user. However, single user does not harm the system at much high level. The research presents the first attempt to identify and validate Sybil groups in RenRen online social network identifying over 2600 groups between 985,000 users by analyzing action time similarity of user in groups (Jiang et al., 2012).

An effective network security strategy requires identifying threats and then choosing the most effective set of tools to fight them. There are three ways to deal with attacks in network that are Detection, Prevention and Counter Measures which are studied by Ankita S. Koleshwar and research team in the article published by international journal of ARCS (Koleshwar et al. 2014).

Kuan Zhang, Xiaohui Liang, Rongxing Lu and Xuemin (Sherman) Shen, researched on Sybil attacks and their defenses in the IoT proposed survey Sybil attack and defense system in IoT. Their research explained about the types of Sybil attacks considering Sybil attacker's capabilities. Also, the research presented some Sybil defense schemes, with a social graph based Sybil detection, behavior classification based Sybil detection and mobile Sybil detection with the comprehensive comparisons (Zhang et al., 2013).

Research paper by Zhang K. and team survey Sybil attacks and defense schemes in IoT. First, define three types Sybil attacks: SA-1, SA-2, and SA-3 according to the Sybil attacker's capabilities then present some Sybil defense schemes, including social graph-based Sybil detection (SGSD), behavior classification-based Sybil detection (BCSD), and mobile Sybil detection with the comprehensive comparisons including discussion of research issues (Zhang et al., 2013).

Research present two forwarding protocols for mobile wireless networks of selfish individuals. Extensive simulations with real traces show that protocols introduce an extremely small overhead in terms of delay, while the techniques introduce in research help to force faithful behavior have the positive and quite surprising side effect to improve performance by reducing the number of replicas and the storage requirements. The protocols also tested in the presence of a natural variation of the notion of selfishness - nodes that are selfish with outsiders and faithful with people from the same community. Even in this case, protocols are shown to be very efficient in detecting possible misbehavior (Mei et al., 2012).

Encounter-based social networks and encounter-based systems link users who share a location at the same time, as opposed to the traditional social network paradigm of linking users who have an offline friendship. This new approach presents challenges that are fundamentally different from those tackled by previous social network designs (Mohaien et al., 2013).

In the research Defense against Sybil Attack in Directed Social Network researcher attempt to solve the problem of defense against sybil attacks in directed social networks. Then proposed a set of measures for the quality of network partitions, with modularity as a special case and presented an algorithm based on the set of measures and iterative optimization to detect the sybil region (Liu P. et al., 2014).

In the research paper by Daniele Quercia and Stephen Hailes (2010), the authors proposed a new decentralized defense for portable devices and called it MobID. The aim of their research is to manage two small networks in which it stores information about the devices it meets it network of friends that is the honest devices, and its networks of foes contain suspicious device. With the help of these two networks, the device is then able to determine whether the unknown individual is carrying out a Sybil attack or not (Quercia and Hailes, 2010).

Mining (Social) Network Graphs to Detect Random Link Attacks research by Nisheeth Shrivastava and team provided a generic abstraction of spam emails, annoying tele-marketing phone calls, viral marketing in social networks attacks, called the Random Link Attack (RLA), that can be used to describe a large class of attacks in communication networks. The research mine the social networking graph extracted from user interactions in the communication network to find RLAs and formally define RLA and show that the problem of finding an RLA is (theory) NP-complete (Shrivastava et al., 2008).

More and more evidences show that some real social networks are not fast-mixing, especially when strong trust relationships are considered. Moreover, the accuracy of all existing solutions is related to the number of attack edges that the adversary can build. Their research proposed the local ranking system for estimating trust level between users. Their scheme has three unique features. One, System is based on both trust and distrust relationship. Second, instated of storing entire social graph, user carry limited information related to themselves and third, their system weakens the impacts of attack edges by removing several suspicious edges with high centrality. They have validated the effectiveness of the system they have designed through comprehensive experiments (Chang et al., 2013).

In Research presented paper SybilDefender, a sybil defense mechanism that leverages the network topologies to defend against sybil attacks in social networks. Based on performing a limited number of random walks within the social graphs, SybilDefender is efficient and scalable to large social networks (Wei et al., 2012).

Research paper Enabling Trustworthy Service Evaluation in Service-Oriented Mobile Social Networks published by Xiaohui Liang and team identify three unique service review attacks in propose a Trustworthy Service Evaluation (TSE) system to enable users to share service reviews inservice-oriented mobile social networks (S-MSNs), i.e., linkability, rejection, and modification attacks, and develop sophisticated security mechanisms for the TSE to deal with these attacks with extension of the bTSE to a Sybil-resisted TSE (SrTSE) to enable the detection of two typical sybil attacks (Liang et al., 2014).

Web transaction data usually convey user task-oriented behaviour pattern. Web usage mining technique is able to capture such informative knowledge about user task pattern from usage data. With the discovered usage pattern information, it is possible to recommend Web user more preferred content or customized presentation according to the derived task preference. The proposed research on a Web recommendation framework is based on discovering task-oriented usage pattern with Probabilistic Latent Semantic Analysis (PLSA) model. Combining the identified task preference of current user with the discovered usage-based Web page categories, research present user more potentially interested or preferred Web content (Xu et al., 2006).

Yan Sun, Lihua Yin, and Wenamao Liu discussed Defending Sybil Attacks specific types of MSNs based on past researchers focused. They proposed a security mechanism to detect Sybil nodes and eliminate them to ensure the routing security while routing forwarding by measuring the distance in client side and eliminating Sybil node in server side. The research also demonstrates the solution is correct and analyzes its energy costs (Sun et al., 2014).

The mobile spam is getting increasingly serious, troubling users daily life and ruining the service quality. The researcher Yuhang Zhao and team proposed a novel approach for spam message detection based on mining the underlying social network of SMS activities. Experimental results on real dataset illustrate effectiveness of various features, showed desired results. With the help of the robust structure of social network constructed from SMS, team able to overcome the frailties of traditional methods such as keywords and flow rate filtering (Zhao et al., 2012).

3 Research Methodology

Graph theory and the newly developed graph system have been widely adopted in the network analysis nowadays. The newly developed graph system explains the outcome through the graph and find the Sybil user in honest community, behavior analytics system which help to trace the behavior of the Sybil user in honest as well as Sybil region, Social Network-based Sybil defense system which contributes to identify Sybil defenses suffer high false positive rates, Neighboring monitoring-based Sybil defense which contribute to identify the real and fake identity in local stored region, Signed social network based system contributes to identify trust and distrust relationship among user but more to identify distrust relationship, performance analysis and evaluation system that also contribute to identify the impact of honest community with the assistance of graphs again and many more.

Objective1 To identify the behavior between nodes on mobile social network by the connectivity they have made

The Sybil entity can recognized by the connectivity between Sybil user and other honest users by using the cross checking method. This could be a number of times of connection performed by a single user or multiple users. Suspicious identity can trace by higher number of connections between individual entities. Most Sybil attackers behave similarity as normal users and share the information, contents with other Sybil numbers of times online. Multiple connections with a single user could help to recognized Sybil user.

Objective 2 Identify specific link between honest region and Sybil region by irregular time of connection

The Sybil entity could be traced in two ways. Firstly, regular relationship with single or multiple users for a longer time. This will help to identify Sybil user. Identifying the connection time between two nodes helps to track the Sybil user under honest region. Secondly, the irregular time of connection between two nodes creates suspicious activity. These intermittent connections could be due to the period, abnormal time of connection, different location and IP address each time, same user connection, no same user connection ever, etc.

Objective 3 Develop the algorithm(s) with graph-based system to identify the connectivity intensity or the number of times of connections between nodes

The research will continue with the help of following methods:

4 Research Evaluation

First, Graph-based System and behavior analytics will produce the express the results using MATLAB and SPSS software system.

Step 1 Generating a graph using MATLAB and SPSS and build a relationship between user 1 and user 2 of the dataset Infocom06 to achieve objective 1

There are many relationships among the users who discover the explicit linkage between each other. The shortest and quickest method to stars the research is to evaluate the dataset through the graphical technique by plotting some random graphs considering the relationship among the users. This task will perform as follows:

Use of a bar graph, consider x-axis as a user one (Column one in dataset) and the y-axis as user 2 (Column 2 in dataset). This chart will show the maximum user connection of user 2 with user on as shown below in Fig. 1. It explores the highest connectivity between users as well as the lowest connectivity between each user. The aim behind to produce this graph is to identify the Sybil nodes that are a number is higher in connectivity with the particular user.

Furthermore, the generated graph also expresses the average user connections with each other which are considered as an honest node at this stage. Because most of honest nodes seem connected with other nodes, those have the same average connections (Fig. 2).

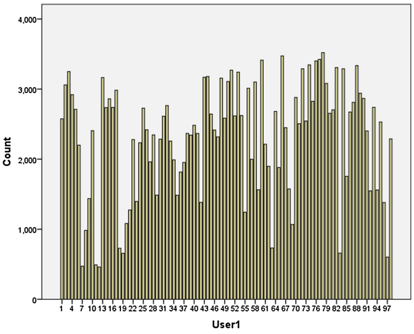


Fig. 1. User 1 vs. User 2

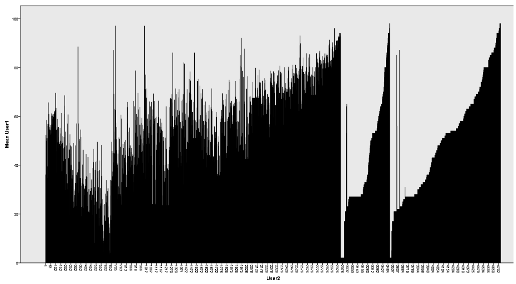


Fig. 2. User 2 vs. User 1 suspicious identity with high connection and low connections

Step 2 Generating a graph using SPSS and produce a relationship between user 2 and user 1 of the dataset infocom06 to achieve objective 1

In this section, the research flips the Step 1 and produce the opposite result to cross-checked the user connectivity. This provides the connectivity of the users to with user 1. This graph gives the crucial outcome. The connectivity with user between each other is very high, and it gets lower and lower until the last user of column2. This graph contributes to the intensity of the user connections.

The algorithm produces both graphs as follows, (Note: Algorithm may get change in further process of research):

Algorithm 1(As Depicted in Fig. 1)

```
Algorithm to plot Attribute ID vs. Quantity (Highest
connection)of user connected to User ID Graph.
```

```
Procedure: Attribute ID vs. Quantity of user
X = Dataset Column 1 // Attribute ID
Y = Dataset Column 2 // Max number of users connected to
Attribute ID
Assign a, b, c as integer
Variable c set to 1
While X <= Xmax do
    Set c to 0
While Y <= Ymax do
      Assign X to a
      Assign Y to b
       Increment X by one// (row =row +1)
      Check condition if (X == a) AND (Y == b) THEN
        Increment X  by one
 Else
    Increment c by one // User Counter
      End if
     End while
PLOT (Dataset, X, c)   // plot single bar on graph
Increment X by one
End While
End Procedure
```

Definition: Maximum number of connection

$$\mathrm{Hn} = \text{Honest entity}$$
$$\mathrm{Sn} = \text{Sybil entity}$$
$$\lim \mathrm{Hn} \to \partial$$
$$0 < \partial < 300$$
$$\lim \mathrm{Sn} \to \partial > 300$$

From the above 3 algorithms, the second figure or graph have contributed which

Algorithm 2 (As Depicted in Fig. 2)

Algorithm to plot User ID vs. Quantity of Attribute ID Graph.

```
Procedure: User ID vs. Quantity of Attribute ID
X = Dataset Column 1
Y = Dataset Column 2
Assign a, b, c as integer
Variable c set to 1
While X <= Xmax do
    Set c to 0
While Y <= Ymax do
      Check condition if (X == a) AND (Y == b) then
        Increment X by one
        Increment c by one // User Counter
End if
If X is not equal to a then
    PLOT (Dataset, Y-axis, c) // plot single bar on graph
    set c to 0
End if
    End while
Increment X by one // X-axis row
End while
End procedure
```

Algorithm 3 (As Depicted in Fig. 5)

Algorithm to plot User ID vs. High short-time connectivity with User ID Graph.

```
Procedure: User ID vs. High short-time connectivity
X = Dataset Row 1, Column 1
Y = Dataset Row1, Column 2
Assign a, b, c as integer
Variable T1 = Dataset Row1, Column3
Variable T2= Dataset Row2, Column 4
While X <= Xmax do    //Outer loop
    Assign X to a
      Assign Y to b
    While Y <= Ymax do // Inner loop
      Check condition If T1 Same as T2 then
        Increment c by one // User Counter
End if
Increment X by one
Variable d set to Y
If b is not equal to d then
PLOT (dataset X, c) // plot single bar on graph
    Set c to 0
End if
    End while
End while
End procedure
```

Definition: Short time connectivity:

$$\lim \mathrm{Hn} \rightarrow \partial$$
$$0 < \partial < 1\, sec$$
$$\lim \mathrm{Sn} \rightarrow \partial$$

will be explained further in this research (Fig. 3).

The other results produce using MATLAB and SPSS as follows:

Table 1. User 1 Vs. Short time Count

Nodes	Connected times
88	1978
77	1901
78	1889
72	1884
82	1883
66	1862
84	1848
89	1846
76	1809
60	1804

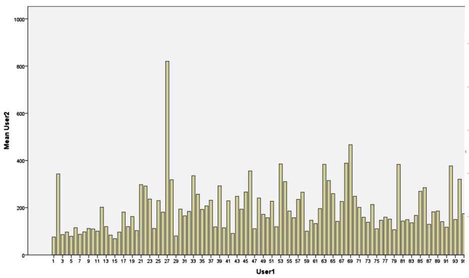


Fig. 3. Highest connecting user 1 vs. user 2

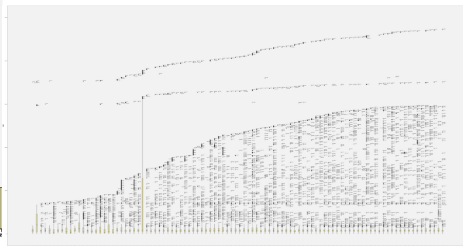


Fig. 4. Boxplot User 1 Vs. User 2 shows the user gap between one to other.

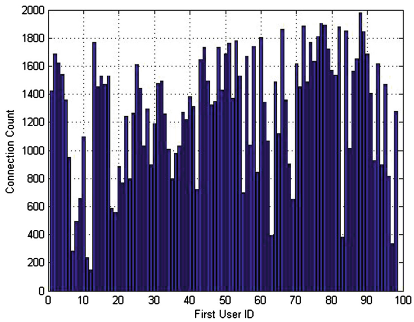


Fig. 5. Bar graph User 1 Vs. High Short time connectivity (Between various Connections).

Table 2. Attribute ID vs. Quantity (Highest connection)

Nodes	Connection count
27	~800
69	~400
2	~350

Table 3 Analyse - descriptive statistics – frequencies (high frequencies repetition and engagements of beginning user of column 1 on by their position in between nodes.)

Node	Frequency (Descending order)	Position in sorted table
3	3250	12
2	3060	21
4	2920	25
5	2711	36
1	2576	46

Table 4. Analyse - descriptive statistics- frequencies (high frequencies repetitionandengagements of beginning user of column 2 on by their position in between nodeschanges dramatically) (~ means continueation of numbers)

Node	Frequency (Descending order)	Position in sorted table
88	3543	1
93	3319	2
~		
85	1392	70
39	1367	71
~		
189	1025	89
63	986	90
~		
212	99	185
154	94	186
3481	10	509
172	9	510
3502	2	3253
15	1	3254
4723	1	4518
4724	1	4519

Table 1 demonstrates the high degree of a short time cumulative count with column 1 user in dataset, produced based on Fig. 5. Table 2 explains the high level of connection of column 1 user engaged in user 2. From the dataset, the users those are from column 1 (considered as user1) got involved in dataset column 2 (considered as user2) shows the high degree of connecting count compare to others. Table 3 explains the Position in sorted table (Descending Order) Analysed descriptive statistics frequency of column 1 user in dataset high frequencies repetition and engagements of beginning user of column 1 on by their position in between nodes (Fig. 4).

Table 4 explains the Position in sorted table (Descending Order) Analysed descriptive statistics frequency of column 2 user in dataset high frequencies repetition and engagements of beginning user of column 2 on by their position in between nodes.

Continuing the Tables 1, 2, and 3 outcome research, produced the Table 4 result that shows the dramatic change in frequency of nodes connection from user 2 to 4724 users of column 2 in data set. The connecting user frequency counting higher to lower, position sorted in the table is significantly 1 to 100 in between frequencies 3543 to 900 (number of connections). And all other nodes in position of frequencies >900 are in significantly positioned of <100 up to 4519 in sorted table. Figure 6 below produced using SPSS explains the Table 4 results.

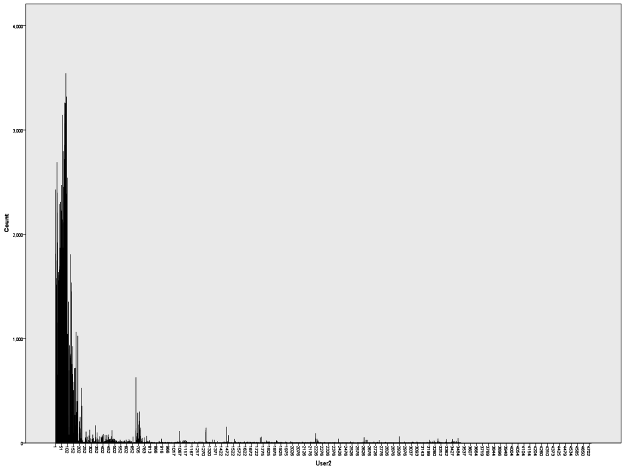


Fig. 6. Bar Diagram for User 2 Vs. High Short time connectivity

5 Conclusion

Content Analysis is the general set of technique that is useful for analyzing and understanding the collection of text. The biggest disadvantage of the content analysis is that, even if it is most accurate, it is very time-consuming and too much manual work is involved in it. The connection analysis is use to identify the connection between two and more nodes in the database that help to discover and understand the establishment and behavior between each connection. It is based on connections, frequencies and its time of contact with amongst the nodes. It may not have most accurate results, but the process to identify Sybil user is quicker. This technique provides the limited workload with necessary information to the administrator to process and identify the nodes quick action.

In the conclusion of this research, it has identified that the Infocom06 dataset containing users has some Sybil user connections among them. The graph-based system helps to evaluate the results of this introductory outcome of this research. The above research findings based on connection analysis, show that the Sybil users among the honest user can be identified and evaluate using graph-based system. The graphs output express that, there are some Sybil users who have a highest connectivity to other trustworthy entities. For example, the user 27th is considered as a highly suspicious as it has highest connection then other users. Further, the research will continue experiments based on connection analysis and content, and it will evaluate research outcome for substantial proof with the few more assumptions using behavior analysis.

The behavior analytics method will be to search the answers with new assumptions. With this continues graph based evaluation, the behavior system will be implemented based on the above finding to trace the behavior of these Sybil node.

References

Alvisi, L., Clement, A., Epasto, A., Lattanzi, S., Panconesi, A.,: SoK: the evolution of sybil defense via social networks : security and privacy (SP), 2013: In: IEEE Symposium on security and privacy, pp. 382–396 (2013)

Jyothi, B.S., Janakiram, D.: SyMon: a practical approach to defend large structured P2P systems against Sybil Attack. Peer-To-Peer Networking and Applications **4**(3), 289–308 (2011)

Cortimiglia, M.N., Renga, F., Ghezzi, A.: Mobile social networking: a case study in an australian mobile network operator, mobile business (ICMB). In: 2011 Tenth International Conference on Mobile Business, pp. 84–92 (2011)

Haifeng, Y., Gibbons, P.B., Kaminsky, M., Feng, X.: SybilLimit: a near-optimal social network defense against sybil attacks. IEEE/ACM Trans. Netw. **18**(3), 885–898 (2010)

Haifeng, Y., Kaminsky, M., Gibbons, P.B., Flaxman, A.D.: SybilGuard: defending against sybil attacks via social networks. IEEE/ACM Netw. **16**(3), 576–589 (2008)

Hao, X., Weidong, X., Tang, D., Tang, J., Wang, Z.: Core-based community evolution in mobile social networks : big data, 2013. In: IEEE International Conference (2013)

Jing, J., Zifei, S., Wenpeng, S., Xiao, W., Yafei, D.: Detecting and validating sybil groups in the wild, distributed computing systems workshops (ICDCSW). In: 2012 32nd International Conference, pp. 127–132 (2012)

Koleshwar, A., Thakare, V., Sherekar, S.: Study of mobile cloud computing security against cyber attacks. Int. J. Adv. Res. Comput. Sci. **5**(4), 187–191 (2014)

Kuan, Z., Xiaohui, L., Rongxing, L., Xuemin, S.: SAFE: a social based updatable filtering protocol with privacy-preserving in mobile social networks, communications (ICC). In: IEEE International Conference, pp. 6045–6049 (2013)

Kuan, Z., Xiaohui, L., Rongxing, L., Xuemin, S.: Sybil attacks and their defenses in the internet of things. Internet Things J. IEEE **1**(5), 372–383 (2014)

Mei, A., Stefa, J.: Give2Get: forwarding in social mobile wireless networks of selfish individuals. IEEE Trans. Dependable Secure Comput. **9**(4), 569–582 (2012)

Mohaien, A., Kune, D.F.: Secure encounter-based mobile social networks: requirements, designs, and tradeoffs. IEEE Trans. Dependable Secure Comput. **10**(6), 380–393 (2013)

Pengfei, L., Xiaohan, W., Che, X., Chen Z., Gu, Y.: Defense against sybil attacks in directed social networks. In: 2014 19th International Conference Digital Signal Processing (DSP) (2014)

Quercia, D., Hailes S.: Sybil attacks against mobile users: friends and foes to the rescue. In: 2010 Proceedings IEEE, INFOCOM (2010)

Shrivastava, N., Majumder, A., Rastogi, R.: Mining (social) network graphs to detect random link attacks. In: IEEE 24th International Conference on Data Engineering, ICDE, pp. 486–495c (2008)

Wei, C., Jie, W., Tan, C.C., Feng, L.: Sybil defenses in mobile social networks,: In: 2013 IEEE Global Communications Conference (GLOBECOM), pp. 641–646 (2013)

Wei, W., Fengyuan, X., Tan, C.C., Qun, L.: SybilDefender: defend against sybil attacks in large social networks, In: 2012 Proceedings IEEE, INFOCOM, pp. 1951–1959(2012)

Xu, G., Zhang, Y., Zhou, X.: Discovering task-oriented usage pattern for web recommendation, In: Sevents Australiasian Database Conference (ADC), (CRPIT), vol. 49 (2006)

Yan, S., Lihua, Y., Wenmao, L.: Defending sybil attacks in mobile social networks. In: 2014 IEEE Conference Computer Communications Workshops (INFOCOM WKSHPS), pp. 163–164 (2014)

Yuhang, Z., Zhaoxiang, Z., Yunhong, W., Jianyun, L.: Robust mobile spamming detection via graph patterns. In: 2012 21st International Conference Pattern Recognition (ICPR), pp. 127–132 (2012)

Zhang, B., Wang, Y., Vasilakos, A.V., Ma, J.: Mobile social networking: reconnect virtual community with physical space. Telecommun. Syst. **54**(1), 91–110 (2013)

Zhang, H., Dantu, R.: Predicting social ties in mobile phone networks, intelligence and security informatics (ISI). In: 2010 IEEE International Conference, pp. 25–30 (2010)

Zhang, H., Dantu, R., Cangussu, J.W.: Socioscope: human relationship and behavior analysis in social networks. IEEE Trans. Syst. Man Cybern. Part A: Syst. Hum. **41**(6), 1122–1143 (2011)

Internet of Things: A Prototype Architecture Using a Raspberry Pi

José Ignacio Rodríguez Molano(✉), Diana Betancourt, and Germán Gómez

Universidad Distrital Francisco José de Caldas, Bogotá, Colombia
jirodriguez@udistrital.edu.co,
dianabetancourtyate@gmail.com, geangoro@outlook.com

Abstract. This article contains a description of Internet of things, the principles on which it is based, the elements and technologies available to achieve communication between people and objects and applications that have been developed in different areas and demonstrating the importance of the implementation of this current. Also there describes a monitoring prototype developed under the frame of the Internet of Things and implemented through the micro computer Raspberry Pi, a cloud storage server and a mobile device.

Keywords: Internet of things · Ubiquitous computing · M2M communications · Applications · Prototype · Monitoring · Raspberry Pi

1 Introduction

The continuous advances in technology has allowed the development of new methods of communication between people and objects, which allow the exchange of information takes place within the parameters of speed and security. Internet of things has taken advantage of these technological advances and the inclusion of new elements within information systems to allow remote access and control of different systems. Under the philosophy of ubiquitous communication and machine to machine communication, internet of things is defined as a set of technologies designed to allow the connection of heterogeneous objects trough different networks and communication methods; its main objective is to position intelligent devices in different locations to capture, store, and manage information to be accessible anywhere in the world for anyone.

2 What Is Internet of Things?

Internet of Things or IoT was proposed and developed by the laboratory network worldwide research in the field of internet of things, Auto-ID Labs, in 1999; IoT is a network based on radio frequency identification, linking objects by sensing devices and Internet [1]. Thus it is possible to characterize in real-time any type of electronic device and/or environmental element.

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 618–631, 2015.
DOI: 10.1007/978-3-319-21009-4_46

Internet of Things as network can touch any object and body through middlemen links. This network has the means available to achieve data collection anytime and transfer information via communication networks for its processing through cloud computing or smart computing [1].

Tan and Wang [2] provide that the Internet of Things is the direct future of computing and communications. For its development is necessary the combination of different and innovative technologies as support; these technologies are defined like [3]:

- Wireless tracking and Technology.
- Technology sensors for detecting elements in the environment.
- Intelligent technologies such as intelligent materials and intelligent networks.
- Miniaturization technologies to reduce objects.

3 Characteristics of Internet of Things

Internet of Things is the result of new technologies and several complementary technical developments that provide capabilities that collectively help to bridge the gap between the virtual and physical world [4]. These capabilities include:

- Communication and cooperation: the objects have the ability to interface with the resources of the internet and even each other to make use of the data, services and update their status; in this measure, the wireless technologies such as GSM and UMTS, Wi-Fi, Bluetooth, ZigBee are highly relevant.
- Addressing capability: the objetcts in Internet of Things can be located and configured remotely.
- Identification: the objects can be uniquely identified through RFID technology (Radio Frequency IDentification), NFC (Near Field Communication) and barcode scanning.
- Perception: the objects may collect information about its environment through sensors that can record this information, send it or react to this.
- Information processing: the smart objects have storage capacity and have a processor that allows you to interpret information.
- Location: the smart things have knowledge of their physical location or can be easily located. This is achieved with technologies such as GPS (Global Positioning System) or the mobile network.
- User Interfaces: the smart objects can properly communicate with people (either directly or indirectly, for example through a smartphone). Here are critical new interaction paradigms as tangible user interfaces, flexible displays based on polymers or methods of speech recognition, pictures or gestures.

Internet of Things proposed connecting all physical objects based on Internet infrastructure to exchange information; in this vision devices and objects are no longer disconnected from the virtual world but that can be remotely controlled and act as points of access to services.

4 Principles of Internet of Things

4.1 Ubiquitous Computing

Ubiquitous computing emerged in early 1990 by Mark Weiser. This concept is defined as a method for improving computer use effectively making it invisible to the user and is mainly characterized by the connection of things to computing [5]; thus the people it can focus solely on the task and not the tool [6].

Ubiquitous computing or UbiComp has two main objectives: (a) reduce the amount of attention that users invest in their devices and (b) create the necessary interfaces to access any information at any time and any place [7].

The concept of ubiquitous computing is a paradigm shift from a traditional view of computing. This new trend of information and communication technologies is based on miniaturization, thanks to that mobile devices and smart devices have become essential but invisible elements in the daily lives of anyone; such devices are equipped with sensors and communication systems which allow them to interact with the environment to collect information and then share it with someone, i.e. ubiquitous computing enables access to all kinds of information for anything anywhere. Currently the emphasis of studies related to ubiquitous computing refers to the practical uses that may develop within different social and work environments [8].

The environments in which is possible apply the Ubicomp should have the following mandatory requirements, which are critical to the success of this new trend based on an everyday environment: (a) constant use of different devices to desktop computers (tablets, Smartphones, smart sensors, roomwares, etc.) and (b) a stable wireless infrastructure to provide a constant flow of information.

It is also possible to classify the application environment of UbiComp depending on the work being performed; There are four general classifications [7].

- Creative environments: generating ideas relating to products or projects, flexible connectivity for input and output devices necessary.
- Meeting Environments: support tools that enable the integration of persons providing the explanation of ideas and viewpoints (electric whiteboards, wireless projectors, etc.).
- Smart environments: gifted everyday spaces of intelligent tools for analysis and observation of the environment and automatically act according to certain preset parameters (dedicated to facilitate user operations).
- Environmental settings: settings are fully integrated with connected wireless networks and allow intelligently control the information required by the user friendly interface devices.

4.2 Machine to Machine Communications

Machine to Machine Communications (M2M) is the combination of information technology and communication with machines to provide a means to interact with each other using minimal human intervention, aiming to increase the comfort and safety of the end user [9].

This Machine to Machine communication refers to technologies that allow systems communicate with other devices of the same features through other devices such as sensors, which allow to capture and transmit event data through to a software application. Regardless of the type of machine or the type of data, information flows generally in the same way: from a machine through the network and conducted through a gateway to a system where it is processed.

In the M2M architecture can distinguish two members: the M2M users, who are the people who are connected to the network via computers or similar equipment, and M2M terminals, which are third generation devices (tablets and smartphones) especially adapted for this type of communication.

Communication between systems and data transfer can be in two ways: uplink, to collect process information and downlink, for sending instructions, software updates, or to remotely control computers [4]. The basic elements that appear in all M2M environments are:

- System to be managed, which can be alarm mechanisms, control systems energy expenditure, information devices, weather stations, among others.
- M2M device connected to the computer terminal that provides communication with the server and handles the interaction with the elements to be monitored.
- Server, which is the computer that manages the sending and receiving of information systems.
- Communication network that handles data transmission either through wires or wirelessly.
- Applications that are responsible for collecting, storing and analyzing information collected by the devices and make the decisions necessary action.

5 Related Technologies

The most relevant technologies associated with the development of the Internet of Things and their applications are RFID (Radio-Frequency Identification), EPC (Electronic Product Code) and NFC (Near Field Communication).

5.1 Radio-Frecuency Identification

Radio Frequency Identification (RFID) is a technology for the complete identification of objects of any type that allows rapid data capture automatically using radio waves [10].

RFID is an automatic identification method based on the storage and retrieval of data [11]. The basic premise behind RFID systems is to mark items with tags. These tags contain transponders (receiving devices and signal transmitters) that emit readable RFID messages. One of the most common extensions of object identification techniques is the tracking function. People can use RFID technology to transfer physical objects to digital.

The identification data entry into the system with additional information such as time, location or sensor data, and together give a new feature to the objetcts: traceability [12]. In addition to traceability, with the use of this technology the following advantages are obtained [10]:

- Increased automation in the process of reading tags since reading can be done without having a direct line of sight with the reader.
- Save time reading the cards as it is possible simultaneous input from more than one label.
- Full visibility of all information stored as information on the label remains intact.

A RFID system is usually composed of three elements: tags, readers and middleware to integrate data with different applications.

RFID Tags. RFID technology uses a microchip cards equipped to store data and a printed circuit as a transmitting antenna, used to communicate via radio frequency signals. RFID tags can be attached to any item and can be adapted to any conditions (humidity, dust, dirt, etc.).

The tags are classified into active and passive categories depending on their source of electricity. Active tags have their own power source and to transmit a stronger signal. The readers can access them from a farther distance (20–100 m), likewise these features make these tags are larger and more expensive. On the other hand, passive tags are cheaper and smaller, these don't have an integrated battery and collect energy from the electromagnetic field created by the reader [13].

Readers. Areader is an electronic device that communicates with tags trough the antenna and reading the information stored in the RFID tag. The reader may have various forms of design, whether in a fixed or moving terminal [14].

RFID reader creates a field of radio frequency waves and detecting and may be able to read data from a transponder and write data to this. When an RFID tag passes through a radio frequency field generated by a compatible reader, the tag reader reflected back to the identification information about the object that has the tag.

Middleware RFID. It is a special type of software that is used to collect and filter data from RFID reader devices. Through this software is managed in real time reading information that made readers, the processed data are collected, transformed and transferred to other existing information systems [14].

5.2 Electronic Product Code

The Electronic Product Code (EPC) is a universal identifier based on Universal Resource Identifiers (URIs); This code provides a unique identity to each physical object anywhere in the world and for all time [15]. The EPC is designed to facilitate processes and applications requiring manipulate data visibility of physical objects. This numbering scheme provides a unique identification for objects and systems; This identification must be large enough to list all objects and to accommodate all methods of allocation of current and future names.

The numbering is based on EPC-64, 96-EPC and EPC-256, three encoding modes, which are respectively 64, 96 and 256 bits long [16]. No matter what type of coding mode is used, the EPC includes:

- Header that identifies the length, type, structure, version and generation of EPC.
- Manager Number that identifies the company or manufacturer of the object.
- Kind of objetc.
- Serial number, which is the specific instance of the class of object tag.

RFID tags store the only EPC in a chip and transmit this code across an antenna for readers of RFID. In an environment of global network, a reader reads the code objects and transmitted to the server object naming (ONS). When viewing the EPC-server information service using the code can obtain detailed information of the object [11].

5.3 Near Field Communication

Near Field Communication (NFC) is a wireless communication technology devices (especially mobile phones and personal digital assistants). This technology was developed by Philips and Sony in 2002 and it combines wireless connectivity technology RFID and interconnection technologies to offer a wireless communication short-range and high frequency between two NFC devices located within 20 cm [17].

The NFC systems consist of two elements: (a) the initiator, which starts and controls the exchange of information and (b) the purpose, which is responsive to the requirement of initiating device.

In an NFC system there are two modes: active and passive. In active mode devices generate their own RF field to transmit data. On the liabilities side, only one of these devices generates the RF field while the other is used to load modulation for data transfers [18].

6 Applications

6.1 Medicine and Healthcare

With the development of the Internet and the emergence of electronic possibilities for diagnosing diseases, there is a modification of the internet of things which allows using wireless sensor networks and Internet to integrate information flow between hospitals, doctors and patients, generating a new medical model with next characteristics:

- Traceability systems that control production processes, delivery, anti-counterfeiting and tracking medical equipment and medicines.
- Identification of products by adhering RFID tag in which information is available, so it is a tool for identifying counterfeit drugs and his constant monitoring in real time

- Management of medical waste, the application of RFID tags allow mapping and location of medical waste generating.
- Identification of patients, which includes medical, medical examinations, treatment records and possible allergies to medicines. Administration medical emergencies: by collecting and storing patient information in RFID cards facilitates the identification of patients and their close family members to share information about the health status of persons accessing critical moments in any hospital.
- Neonatal theft system that integrates the identification and collection of information from parents and baby.
- Patient status information stored in smart tags accompanied by alarms that are triggered when some variation exists in its current state [19].

6.2 Logistic

A Logistic Mobile Application (ALMA) is a mobile application that allows taking into account the dynamic nature of the logistical problems by improving service quality and customer satisfaction. This application is based on High Performance Computing or HPC, an infrastructure that makes use of clusters, grids and P2P networks, through a location that has the availability of necessary devices.

In ALMA all products are identified by smart tags; when a product has been delivered, the person in charge of transport scans the tag and transmits the information to the logistics center, this is done through a smartphone connected via 3G or Wi Fi. Similarly anything new that is generated within the transport process is informed through the application. The information is collected by the control center which allow to have a continuous record of the status of any product being transported or to be transported. The ALMA application combines an estructura of communication and activities of computation distributed in order to obtain rapidly solutions for the different disadvantages generated inside the distribution routes or the logistic activities [20].

6.3 System for Food Quality

Due to the importance of the quality of the food inside the process of production and distribution of these, there has appeared a system of summary of information based on the Internet of the things where it is possible to exchange real time information, to share resources and to improve the synchrony between activities. The users of the system of vigilance provide different structures depending on the type of activity that they realize. This application is based in:

- Sensing network is the source of data is not only used for data collection occurring within the manufacturing, processing, distribution and sales but also provides all the tools for data collection, transmission and displaying information this. Devices for gathering information include RFID devices, bar code, QR code readers, video cameras, infrared sensors and other devices.

- Sensing network used for data collection occurring within the manufacturing, processing, distribution and sales.
- The network and transport layer composed by the mobile communication network, internet and other networks of private use.
- The cloud, that allows storage and data management.
- The management layer and application layer designed to ensure the exchange of information between different departments responsible for logistics activities and users.

The RFID tags used inside the manufacturing process are used to register the information of raw materials and manufacturers (traffic of the products, information about the type of storage, etc.), all the compiled information is administered by means of a RFID middleware which takes charge generating the necessary connections to take all the information up to the cap of application that is used by the user [21].

6.4 Telecommunications

The creation of Internet of the things has allowed to extend and to fuse the different technologies and create new possibilities and mass media between the users and the things (GSM, technology NFC, Bluetooth, SIM cards, etc.). For all types of mobile phone communication IoT is a essential part of the system of interaction since it is the reader all the information and allows transfer this information to servers that can be stored, filtered and administered according to user needs.

In the long term and thanks to the development of new possibilities for information transfer of information and the rise of the mobile phones, the border of the Internet of Things and telecommunications will be smaller, generating total situational communication that allowing the users should have an unlimited communication taking advantage of the communication with the environment and the computer science [22].

7 Why Raspberry Pi?

The capture system is based on the Raspberry Pi which is equipped with different sensors capable of detecting changes in the environment and capture the information required by the user. Although there are different types of microcomputers and micro components that could be used in the construction of the base of the capture system, such as Arduino, Banana Pi, Pc Duino, UDOO, etc. The Raspberry Pi compared with Arduino has characteristics that make it a top plate in its functionality as it has hardware that allows: (1) fast data transfer, (2) ports for external hardware (easy connection), (3) internet via Ethernet or Wi-Fi cable and (4) video output (HDMI) and audio (mini-jack); in the market there microcomputers that have the same characteristics as the Raspberry (even improve certain aspects RAM) but these options have a higher purchase price to the Raspberry, positioning it as the best option for the authors to develop this capture system.

8 Prototype Architecture Based on Internet of Things and Using Raspberry Pi

Based on the principles mentioned above a prototype monitoring system is developed on the concepts of ubiquitous computing and M2M communication and implemented under the framework of the internet of things. The aim of this prototype is to capture images of a place and allow access to these real-time and remotely. A environment-object and object-object communication to achieve the proposed aim was established.

The communication environment-object of the system of vigilance allowed to monitor different environmental variables status changes were shown through images, for this there were in use four sensors that allowed determining the flow of people in a place, the gas presence in the atmosphere, surface stability and existence of contact between a person and the prototype. The reading of changes in the state of the sensors made by was processed in a micro computer Raspberry Pi.

Object-object communication between the Raspberry Pi and a mobile device, in order to establish synchronization of images directly to the user across his Smartphone or tablet was established. This communication was achieved through the connection established between the Raspberry Pi and cloud storage server, Dropbox (It is possible to use any server that has within its service tools for application development or SDK); link possible by the ability of micro computer connection wireless internet network.

The monitoring prototype has the following characteristics:

- Using sensor technologies: monitoring prototype uses different sensors for the detection of “weird” elements in the environment and initiate the capture and transmission of images. A sensor acts as a link between what happens in one place and so the user can see this place.
- Collection of data at any time: the prototype does not restrict capture of images and always be aware of every movement present at the place of installation of the system.
- Images transfer: it is possible transfer images within the communication network composed by the user (via a mobile device) and physical system (Raspberry Pi) by means of computing in the cloud (Dropbox).
- User interface: the prototype can communicate with the user through platforms known as personal email (Hotmail, Gmail, Outlook, etc.) or server storage (Dropbox, Amazon, Google Drive, etc.).
- Functional independence: under the guidance of ubiquitous computing reducing the users’ attention on the device, the prototype works automatically and independently of the user, since their capture images and transfer functions are carried out only with the activation of a sensor and a wireless network.
- Communication with the mobile device: the system does not require direct intervention in the hardware or software to operate. The (image transfer) communication conducted under the principles of machine- machine communication in which the micro computer Raspberry Pi and sensors represent the prototype and the mobile device is the terminal which enables communication with the user through a server

and a communication network. Captured images are derived from one machine to another in the same way regardless of the sensor is in use.

- Total disposition of images: the system generates a communication network that allows the access to the user any place and consults at any time of the images captured thanks to the connection of the physical prototype (Raspberry Pi) with the servant online.

Prototype architecture is based on three levels. Each level represents a process to be carried out for that the system meets its objective which corresponds to the capture and synchronization information so that it can be used and/or accessed by the user.

The first level corresponding to External Level, so named because it is the level that is exposed to changes generated by exogenous factors to the system, can also be considered activators system functions (corresponding to some sensors that capture changes in the characteristics environment, each of these detects a specified type of modification of the environment), these activators to detect any alteration by wires send an electrical impulse which is directed to the Raspberry Pi.

After the pulse is generated if there isn't any process to transform it simply be lost, i.e. not triggered any reaction within the system, therefore the Intermediate Level is so important, because allow information flow using the internet and protocols for synchronization of information (SMTP and SSL) to the user which is the Target Level, where mobile devices that can be used to access this information is also included, thanks to the Raspberry Pi is possible to create a logic programming which converts the electrical signal into two types of data, it is possible to generate any reaction due to hardware that has the Raspberry Pi and various output devices that can be found (Figs. 1, 2 and 3).

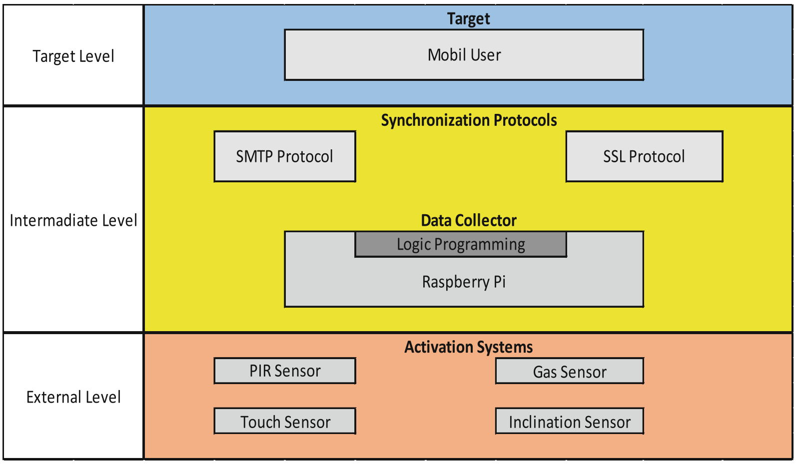


Fig. 1. Prototype architecture based on IoT. Source: Developed by the authors.

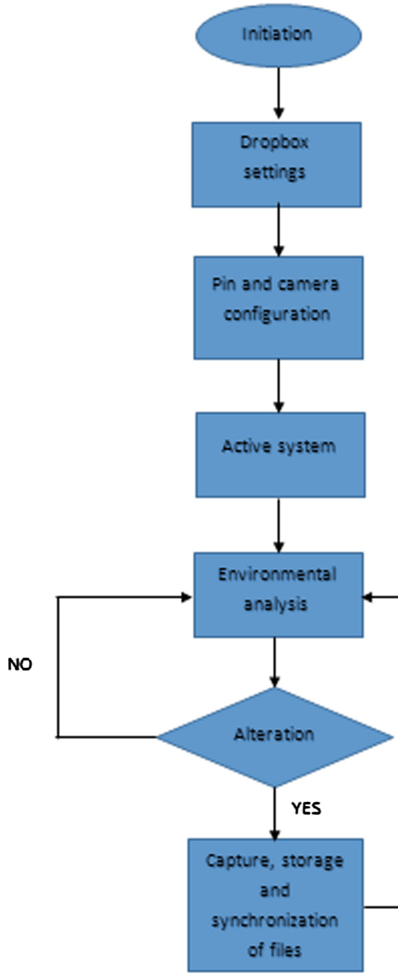


Fig. 2. Prototype flowchart. Source: Developed by the authors.

When executing logic programming are generated procedures that allow the operation of the prototype, from the import of the libraries to capture images if some kind of alteration is detected; processes and the results obtained by executing logic programming are shown below.

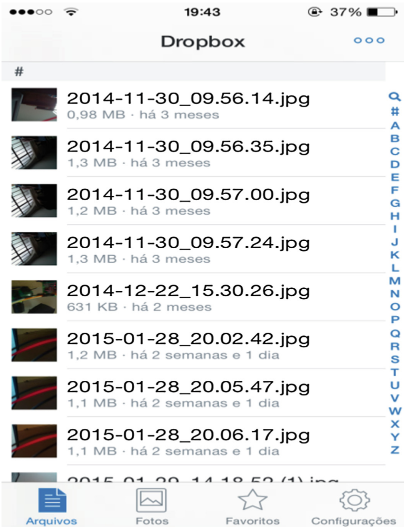


Fig. 3. File synchronization. Source: Developed by the authors.

9 Obstacles and Learning

The development of the system but complex due to little prior contact with a Raspberry Pi but was developable thanks to the possibility of selection a friendly operating system which would be installed within the micro computer, this allowed to perform a exhaustive search for the types of languages and logics of programming used in each of the operating systems Raspberry Pi to pick one.

The next step leads to the identification of digital storage servers which have the tools for application development on Raspberry Pi, multiple options highlighted on the Web (Amazon, Google Drive, Dropbox, OneDrive, etc.), in some of these servers had service cost and other documentation was not sufficiently explicit to support learning and development of the application (remembering the little contact with the Raspberry Pi) was finally chosen Dropbox.

The most simple anf fun: the prototype application scenarios; endless possible uses thanks to the quantity and variety of sensors on the market, each of these focused on analyzing different variables in any environment, allowing the collection of information and analysis of this; the outputs of the Raspberry Pi not only allow you to enable/disable a security system but entire systems as a reaction to an alteration in the environment, this could drastically impact on the decrease of costs and more importantly in safety procedures for human resource but these are theoretical assumptions, the difficult access to active production plants relegated protipo the implementation of an administrative section where environmental changes are few (evidenced by the gas sensor MQ −2) and permiter not reveal the full potential of a small part of the internet of things.

10 Conclusions

The internet of things proposed to establish networks of communication between objects and people that allow bidirectional information flow with minimal intervention on the network. Communication networks focus on direct and global data transmission.

The developed prototype is a communication network which captures, processes and transfers data (images) to the user; this communication network established by connecting a capture subsystem and synchronization subsystem (private information storage server).

Communication between capture subsystem and the mobile device is established through the mobile application storage service Dropbox cloud, created on its website. Through programming language developed the communication link from the Raspberry Pi and Dropbox was created (there are also alternatives like Gmail and Amazon Virtual Private Cloud).

In developing an overall architecture for the analysis of environmental conditions is possible to apply this system to industrial (any process of production and marketing) and the everyday life, thus improving the conditions and ensuring better results in any activity undertaken by the availability of information.

References

1. Guo, L.G., Huang, Y.R., Cai, J., Qu, L.G.: Investigation of architecture, key technology and application strategy for the Internet of Things. IEEE Conf. Publ. **2**, 1196–1199 (2011)
2. Tan, L., Wang, N.: Future internet: the Internet of Things. IEEE Conf. Publ. **5**, V-376–V-380 (2010)
3. Ramirez, G.A.: Evaluación de introducción de internet de objetos en espacios de aprendizaje (2010)
4. Garcia, A.: El Internet de las Cosas y los nuevos riesgos para la privacidad (Tesis de maestria), (2012)
5. Liu, R., Wang, Y., Yang, H., Pan, W.: An evolutionary system development approach in a pervasive computing environment. IEEE Conf. Publ. 194–199 (2004)
6. Liu, Y., Feng, L.: PCA : a reference architecture for pervasive computing. IEEE Conf. Publ. 99–103 (2006)
7. Lupiana, D., O'Driscoll, C., Mtenzi, F.: Taxonomy for ubiquitous computing environments. IEEE Conf. Publ. 469–475 (2009)
8. Sakamura, K., Koshizuka, N.: Ubiquitous computing technologies for ubiquitous learning. IEEE Conf. Publ. 11–20 (2005)
9. Meddeb, M., Alaya, M.B., Monteil, T., Dhraief, A., Drira, K.: M2M platform with autonomic device management service. Procedia Comput. Sci. **32**, 1063–1070 (2014)
10. Miquel, S., Parra, F., Lhermie, C., Miquel, M.: Distribución comercial (2006)
11. Lee, S.-D., Shin, M.-K., Kim, H.-J.: EPC vs. IPv6 mapping mechanism. IEEE Conf. Publ. **2**, 1243–1245 (2007)
12. Weinstein, R.: RFID : A technical overview and its application to technology. IEEE J. Mag. **7**, 27–33 (2005)
13. Rubio, J., Villarroel, S.: Gestión de pedidos y stock (2012)

14. Fuchsíková, P., Kebo, V., Pavel, S.: RFID technology for special food moulds. IEEE Conf. Publ. 183–188 (2012)
15. Karakostas, B.: A DNS architecture for the Internet of Things: a case study in transport logistics. Procedia Comput. Sci. **19**, 594–601 (2013)
16. Liang, Z., Lian, X.: Application research of EPC network in reverse logistics. IEEE Conf. Publ. 369–372 (2012)
17. Sarma, S., Brock, D., Engels, D.: Radio frecuency identification and the electronic product code. IEEE J. Mag. **21**, 50–54 (2001)
18. Chavira, G., Nava, S.W., Hervas, R., Bravo, J., Sanchez, C.: Combining RFID and NFC Technologies in an AmI Conference Scenario. IEEE Conf. Publ. 165–172 (2007)
19. Hu, F., Xie, D., Shen, S.: On the application of the Internet of Things in the field of medical and health care. IEEE Conf. Publ. (2013)
20. El-Baz, D., Bourgeois, J., Saadi, T., Bassi, A.: ALMA, a logistic mobile application based on Internet of Things. IEEE Conf. Publ. 355–358 (2013)
21. Ying, F., Fengquan, L.: Application of Internet of Things to the monitoring system for food quality safety. IEEE Conf. Publ. 296–298 (2013)
22. De Saint-exupery, A.: Internet of Things: strategic research roadmap. Eur. Res. Clust. Internet Things **50** (2009)

Internet of Things for Health: Japanese Consumers' Needs for Preventive Healthcare Products

Yoko Ishino(✉)

Graduate School of Innovation and Technology Management,
Yamaguchi University, Ube 7558611, Japan
ishino.y@yamaguchi-u.ac.jp

Abstract. The Internet of Things (IoT) plays a significant role in a broad range of healthcare applications. However, so far this effort is mainly restricted to actual patients actively undergoing treatment. From the perspective of IoT, these kinds of efforts should spread to the preventive activity of healthy people in the not-so-distant future. Healthy people, whose population is much larger than patients, presumably have various latent needs or wants for an IoT-driven healthcare system, compared with the actual patients. Discerning the real needs from the latent needs of healthy people for preventive care is the key to build up an innovative and favorable IoT-based system. This research tackled this problem from the product device side. We utilized multiple analyses including cluster analysis, graphical modeling, and text-mining, in order to extract from survey data real and useful consumer needs, which can be expected to propel the innovation of health-related home electronics. The effectiveness of the proposed method was verified by a case study using practical questionnaire data, while selecting a blood-pressure monitor and a sleep monitor as case study examples.

Keywords: Knowledge extraction · Healthcare product · Bayesian network · Feature selection

1 Introduction

The healthcare market has been steadily growing in many countries including Japan. In Japan, home electronics relating to health are classified into two types: products within a scope of the Pharmaceutical Affairs Law and other products not covered by the law. With respect to the former, the domestic market of medical equipment has grown to approximately 2.3 trillion yen in 2010, according to the statistics published by Ministry of Health, Labour and Welfare. The factors that seem to contribute to the growth of this field are listed below.

- Japan is becoming a super-aging society, and interest in health and preventive care is increasing further.
- A special medical checkup based on a special health guideline system that was introduced in 2008 has now become a mandatory service. As a result, awareness of metabolic syndrome and lifestyle-related disease prevention, as well as self-health management is increasing among people.

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 632–645, 2015.
DOI: 10.1007/978-3-319-21009-4_47

– Active use of information-communication technology (ICT) has made it possible to commercialize mobile healthcare services (MHS) in a short period, even without a large initial investment. MHS using flat rate packages of a mobile phone are recently attracting attention in Japan, although not yet broadly pervasive.

The last factor is very important, because it is expected that advancements in ICT will open up potentials for telemedicine, home medical care, and new medical solutions. This leads exactly to the Internet of Things (IoT) for health.

In IoT, devices gather and share information directly with each other and a cloud server, making it possible to collect, record, and analyze new data streams faster and more accurately. IoT should provide desirable services especially in the field of healthcare, where its principles are already being applied to improve access to medical care, increase the quality of care, and reduce the cost of care.

IoT-driven healthcare systems are based on the network of sensor devices that connect to the Internet to capture and share human vital data through a secure service layer that connects to a central server. The IoT plays a significant role in a broad range of healthcare applications. Here are some examples of how its potential is already playing out:

- Clinical care: The physiological status of hospitalized patients can be constantly monitored using the IoT-driven and non-invasive sensor devices.
- Remote monitoring: When patients suffering from chronic diseases neither have easy access to hospitals nor present acute symptoms, wireless solutions connected through the IoT can constantly monitor the fluctuations in symptoms.
- Early prevention: Healthy and active people can also benefit from IoT-driven monitoring of their daily activities, as the IoT-driven system can detect some sign of change in everyday activity and report it.

An example of the remote monitoring is MedicalLINK® produced by OMRON Corporation Limited [1], one of the world's biggest medical equipment makers. MedicalLINK® collects automatically acquired data of self-tested blood pressure readings and sends the data to doctors through wireless access to the Internet, such as a 3G network. Moreover, since MedicalLINK® automatically monitors the blood pressure during sleep, doctors can detect circadian variations in blood pressure. Doctors can prescribe proper medication to patients based on the information provided. As of early 2013, 1500 hospitals and medical facilities across Japan had started to use the MedicalLINK® system. This is an innovative and favorable effort, since currently 40 million Japanese suffer from high blood pressure and 16 million of them receive treatment from doctors for the condition. However, so far this effort is restricted to the hypertension patients. In light of IoT, this kind of effort should spread out to the preventive activity of healthy people or potential patients in the not-so-distant future.

Approximately 60 % of Japanese die from cancer, heart disease, or stroke. Improving people's life style before becoming sick can extend their healthy life. Since not only Japan but other developed countries are faced with problems of an aging society, the demand for IoT-driven healthcare systems has been increasing in those countries, and the market of IoT-based systems for preventive medicine has an especially huge potential. Although there is a great chance for an innovative IoT-based system to open up the preventive care market, it is not easy. Healthy people, potential patients whose population is much larger than patients, supposedly have various latent needs or wants within an IoT-driven healthcare system, in contrast to actual active patients whose main need is to recover from disease. In terms of both a product device and a business model, discerning the real needs from the latent needs of healthy people for preventive care is the key to build up an innovative and favorable IoT-based system.

We tackled this problem from the product device side, not the business model side. This research focuses on extracting real consumer needs for pushing innovation in health-related home electronics devices which can ultimately contribute to IoT-driven preventive care. This study used data from a related existing survey that mainly inquired about the penetration level of health-related home electronics, though this research sought not just to accumulate simple statistics on the acquired survey data, but rather to analyze the data's latent relationships and meanings in a much deeper manner. This research executed multiple analyses including cluster analysis, graphical modeling, and text-mining, and then these results were integrated to obtain a profound understanding of consumer needs. Integrating multiple analyses is required to elucidate the real needs, since the thoughts of people generally are complicated and inconsistent. As a result, we obtained detailed knowledge for future innovation directions of health-related home electronics products, which then coupled with the IoT paradigm, will lead to increasingly seamless and further personalized innovative healthcare and health management.

2 Knowledge Extractions from Consumer Survey Data

Using a survey performed to understand the health awareness of general consumers and diffusion of electronic healthcare products, this research found new directions for product development of health-related home electronics on the basis of consumer needs, which is expected to lead to a favorable IoT-driven preventive healthcare system. The outline of the survey used in this research is mentioned below.

The survey regarding home electronics for healthcare was performed through the Internet from Apr. 1 to Apr. 5 in 2014 by MyVoice Communications, Inc. The respondents to the survey were average Japanese consumers who were randomly screened from the panels of this research firm. The number of the valid response was 11334, as the allocation of respondents in terms of sex and age is shown in Table 1.

This study analyzed questionnaire data using several methods in a stepwise manner. The first step was ordinary statistical analyses including cluster analysis, one type of multivariable analysis. The second step is graphical modeling in which subjective feature selection and Bayesian network analysis are used. The third step is text-mining

Table 1. Allocation of respondents in the survey

	Sex		Age				Total
	Male	Female	20's or younger	30's	40's	50's or older	
Number	5538	5796	1318	2394	3439	4183	11334
Ratio	49%	51%	12%	21%	30%	37%	100%

to analyze text data from open-ended questions. Finally, all of these results are integrated to understand real, practical consumer needs.

Of these steps, the second step (graphical modeling) is particularly original, which had been proposed by the author to analyze questionnaire data while solving the feature selection problem and treating multiple objective variables in one model [2]. The outline of the graphical modeling is as follows: (1) an analyst eliminates redundant product features using Cramer's Coefficient of Association (Cramer's *V*) as an index of association between the features, which would be taken account of when consumers choose health-related home electronics, (2) the dependencies between selected product features are inferred using Bayesian networks as consumers' behavior when choosing home electronics, and (3) probabilistic inference is performed to investigate which features have a maximal effect on a specified health-related home electronics. This process and other elemental (consider "fundamental") analytical methods are explained in the next chapter in detail.

3 Elemental Analytical Methods Used

3.1 Cluster Analysis

Cluster analysis is to group a set of objects in such a way that objects in the same group (called a cluster) are more similar to each other than to those in other groups [3]. Although the term cluster analysis itself does not identify a particular statistical method or model [4], there are two typical procedures that can be used to cluster data: hierarchical clustering and non-hierarchical clustering. The typical algorithm of the latter is *k*-means clustering. This research used the former, hierarchical clustering, since *k*-means clustering must have a determined value of the hyperparameter *k* that is the number of clusters before starting analysis.

Hierarchical clustering is based on the idea that an object is more related to nearby objects than to objects far away. Clustering algorithms connect objects to form clusters based on their distance. The results of hierarchical clustering are usually presented in a dendrogram. There are several general distance metrics (also termed distance functions) that define a distance between each pair of objects in the input set. Examples of distance metrics are Euclidean distance, Manhattan distance, Mahalanobis distance, and so on. The choice of an appropriate metric will influence the shape of the clusters. An appropriate metric is usually chosen through a trial and error process. One also needs to decide on the linkage criterion to use, which determines how the distance metric is applied between an object and a cluster, or a pair of clusters. Popular choices are single-linkage clustering (the minimum of object distances), complete linkage clustering (the maximum of object distances), UPGMA ("Unweighted Pair Group Method with Arithmetic Mean", also known as average linkage clustering), or Ward's method (the minimum of the total within-cluster variance). Furthermore, hierarchical clustering can be agglomerative (starting with single elements and aggregating them into clusters) or divisive (starting with the complete data set and dividing it into partitions).

In this research, based on the answer of respondents to the question about their intent to use health-related home electronics, the correlation coefficient matrix of the home electronics was computed. Then, Ward's hierarchical agglomerative clustering method was performed, using the value of the correlation coefficient as the object pair similarity.

3.2 Graphical Modeling Using Cramer's *V* Feature Selection and Bayesian Network

Bayesian Networks. A Bayesian network is a machine learning technique for empirically identifying associations in complex and high-dimensional data. A Bayesian network graphically models probabilistic relationships among a set of variables. Over the last decade, Bayesian networks have become a popular representation for encoding uncertain expert knowledge in expert systems [5]. More recently, researchers have developed methods for deriving Bayesian networks from data [6–11]. When used in conjunction with statistical techniques, a Bayesian network has several advantages for data analysis. The greatest advantage is that a Bayesian network can be used to determine causal relationships from data and, hence, can be used to gain an understanding about a problem domain and to predict the consequences of intervention.

A Bayesian network consists of nodes, edges, and conditional probability tables (CPT). Nodes and edges represent events and conditional dependencies, respectively. In this paper, nodes represent discrete variables, which are questionnaire items related to consumers' attitude and behavior. Since these variables have causal relationships, edges in this paper are directed arrows. This paper considers only a directed acyclic graph, meaning that no loops can occur in the group structure, since such a graph is mathematically manageable.

A Bayesian network can represent nonlinear covariation between features and does not assume a Gaussian distribution of variables, since the relation between features is given in the CPT. The model's structure can be learned from the data based on information criteria such as Akaike's Information Criterion (AIC) and Minimum Description Length (MDL) [12]. In addition, prior knowledge of a model designer can be embedded into the model in advance, as model constraints. The Bayesian network structure obtained from learning enables us to compute the posterior distribution of variables when other variables (the evidence variables) are observed. This process is called probabilistic inference, in which some efficient computational algorithms have been proposed, such as loopy belief propagation [13, 14]. Because of these convenient features, Bayesian networks have recently begun to be applied to express complex human behavior in marketing [15].

Cramer's Coefficient of Association (Cramer's *V*). Although a Bayesian network is a powerful tool, it provides an incorrect structure as a consequence of the structure learning process when the dataset contains too many attributes that are highly dependent on each other. This problem often occurs when a Bayesian network is used to describe complex human behavior, as in marketing area, when similar features are used as explanatory variables.

To overcome the problem, the current study uses Cramer's *V* as a feature selection criterion before starting to learn a Bayesian network structure from the data. Cramer's *V* is a popular measure of association between two nominal variables.

Before using Cramer's V, Pearson's chi-squared statistic should be explained, since Cramer's V is based on the Pearson chi-squared statistic [16]. A contingency table is useful when examining the association between two nominal variables, which relates to a CPT. Based on the contingency table, Pearson's chi-squared statistic is usually calculated to test the independence of two variables. The chi-squared statistic is defined as

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

where O_i is the observed number of cases in category i, and E_i is the expected number of cases in category i. Obviously the chi-squared statistic is affected by sample size.

On the other hand, although Cramer's V is based on the chi-squared statistic, Cramer's V is not affected by sample size and is normalized to the range between 0 (corresponding to no association between the variables) and 1 (perfect association). Cramer's V is given by the following formula:

$$V = \sqrt{\frac{\chi^2}{n(k-1)}}$$

where χ^2 is Pearson's chi-squared statistic, n is the total number of samples (observations), and k is the smaller value among the number of rows and columns.

Calculating the value of Cramer's V for all combinations of two variables would help find redundant variables that have a high association with other variables. This enables us to discard the redundant variables, before starting to learn the Bayesian network structure from the questionnaire data.

Graphical Modeling Using Cramer's *V* Feature Selection and a Bayesian Network. The procedure of the proposed approach is as follows.

Step 1: Calculate the value of Cramer's V for all combinations of two variables

Step 2: Select variables that should be eliminated from the dataset

2-1: The variable combinations in which Cramer's V value is higher than a threshold (decided beforehand, e.g., more than 0.5) automatically denote the candidates to be considered

2-2: An analyst who has domain knowledge of the problem selects variables to be eliminated by examining each value of Cramer's V while focusing on the essential significance of each variable in the relevant field. When the analyst decides that the variables in a candidate combination are semantically redundant, one of them is eliminated. When the analyst decides that the variables in a candidate combination are both important, they are left in the dataset. This process is repeated until there are no further candidate combinations to be considered

Step 3: Learn a Bayesian network structure using the remaining variables

Step 4: Extract useful knowledge from the data by applying probabilistic inference to the obtained Bayesian network structure

3.3 Text Mining

The purpose of text mining is to process unstructured (textual) information, extract meaningful numeric indices from the text, and thus, make the information contained in the text accessible to various algorithms of data mining [17]. In marketing, it is common to include various open-ended questions pertaining to the topic under investigation. The idea is to permit respondents to express their opinions or views without constraining them to particular dimensions or a particular response format. This may yield insights into customers' views and opinions that might otherwise not be discovered when relying solely on structured questions in the questionnaire. For example, you may discover a certain set of words or terms that are commonly used by respondents to describe the pros and cons of a product or service, suggesting common misconceptions regarding the items in the study.

In the text mining process, natural language processing is generally applied as a first step, such as part of speech tagging, syntactic parsing, and other types of linguistic analysis. Then, all words found in the input documents will be indexed and counted in order to compute a table of documents and words, *i.e.*, a matrix of frequencies that enumerates the number of times that each word occurs in each document. Once a table of terms by documents has been derived, all standard statistical and data mining techniques can be applied to derive dimensions or clusters of words or documents, or to identify important terms.

In this research, the term frequency (*TF*) was calculated as a numerical feature value of each term (word). This is identified as follows:

$$TF(w_i) = \sum_j \frac{n_{i,j}}{\sum_k n_{k,j}}$$

where w_i represents the word i, and $n_{i,j}$ is the number of appearances of the word i in the document (a respondent's answer) j. Therefore, $TF(w_i)$ indicates the appearance frequency of the word i. Then, dependency parsing was performed for a document (a respondent's answer). Dependency parsing is one form of syntactic parsing, based on the dependency grammar implicit in sentences. Dependency parsing enabled us to find a certain set of words that were commonly used by respondents to describe desirable health-related home electronics.

4 Results and Discussion

4.1 Fundamental Statistical Results

Figure 1 shows how much the respondents cared about their own health. Consciousness toward health was measured on a 5-point scale: 1 = very careless, 5 = very careful. Around 60 % of respondents cared about their health conditions in a positive way.

Both the current use rate of healthcare electronics and the rate of intention to use them in the future are indicated in Fig. 2. For example, in the case of blood-pressure monitors, the current user rate was 22.9 % and the percentage of respondents who want to use them in the future was 24.3 %, a slight increase. In contrast, the percentage of

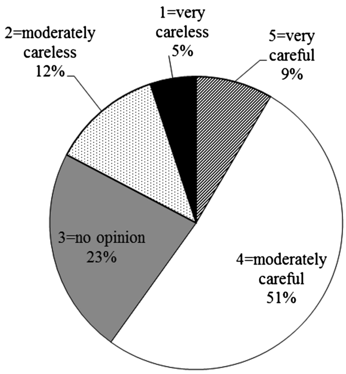


Fig. 1. Health awareness

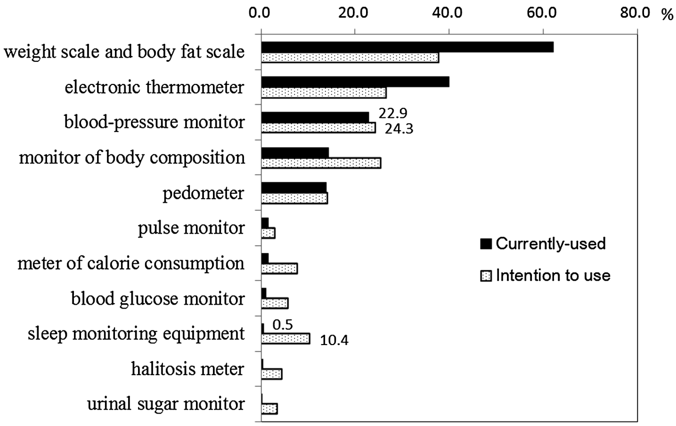


Fig. 2. Use of electronic healthcare products

respondents who want to use sleep monitors greatly exceeded the current user ratio. It jumped up from 0.5 % to 10.4 %, a very remarkable increase.

Figure 3 indicates product features about which respondents consider in choosing healthcare electronics. More than 60 % of respondents take account of the price. We cannot find product-specific trends from Fig. 3, since answers about the products that each respondent wants to use were merged into one result.

4.2 Cluster Analysis

Based on the answer of respondents about their future intention to use health-related home electronics, the correlation coefficient matrix of the home electronics was computed. Then, Ward's hierarchical agglomerative clustering method was performed, using a distance based on the correlation coefficient. The distance is identified as follows:

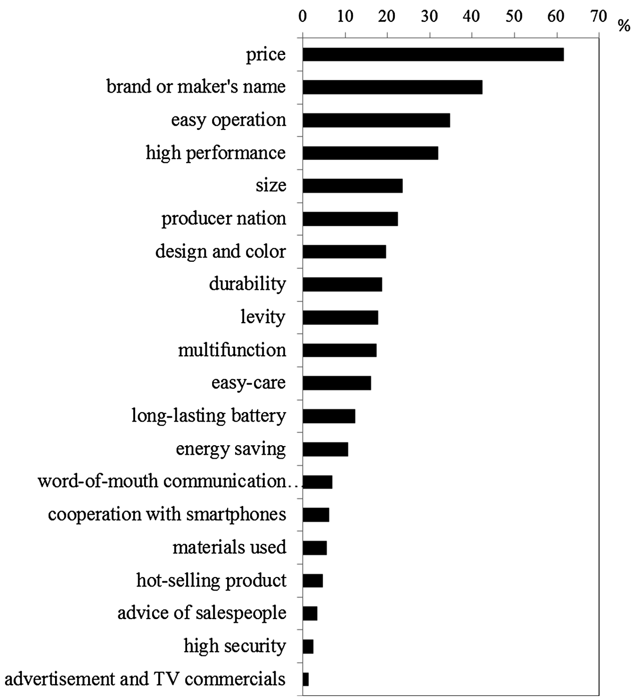


Fig. 3. Product features to be considered in choosing home electronics

$$D_{i,j} = \sqrt{2 - 2 \times C_{i,j}}$$

where $D_{i,j}$ represents the distance between the products i and j, and $C_{i,j}$ represents the correlation coefficient between the products i and j.

Figure 4 shows the result of the hierarchical clustering. The fifteen electronic healthcare products can be divided into 5 clusters. An electronic thermometer and a weight scale formed one cluster. They share the feature that their current use rate is very high. A bad breath sensor and an alcohol sensor formed another cluster, and on separately, a urinal sugar meter and a blood sugar meter also formed a cluster. The former clearly is a breath sensor cluster, and the latter cluster likely relates to diabetes. Although the result seems to be reasonable, other analyses are needed to gain an even deeper understanding of the difference in consumers' behavior or consciousness regarding health-related products.

4.3 Graphical Modeling

From this point, a blood-pressure monitor and a sleep monitor were selected as samples in order to learn about consumers' behavior. Both products increased the future intention of use. However, they belong to different clusters, so that different needs are expected.

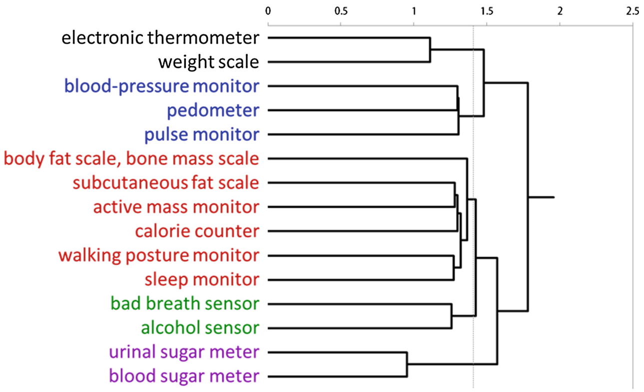


Fig. 4. Clusters of health-related home electronics

The process of graphical modeling was executed using the domain knowledge of an analyst, simultaneously referring to Cramer's *V* values between features as an index. Then, the remaining features were considered as stochastic variables constituting a Bayesian network, and the network structure representing the consumers' behavioral principles was inferred. The software package used was BayoNet version 5.0, where the Greedy Search Algorithm was used as the search algorithm and AIC and MDL were used as information criteria. After obtaining the network structure, probabilistic inference was performed to extract consumer behavior knowledge. Subsequently, text mining was applied to answers of an open-ended question about health-related home electronics that a respondent would want to have in the future. The software package used was Text Mining Studio version 5.0. Finally, those results were integrated and interpreted into knowledge for future innovation directions in health-related home electronics.

Structure of Bayesian Network Obtained. Figure 5 illustrates a best model obtained, in which a sleep monitor and a blood-pressure monitor were adopted as typical healthcare electronics. Four demographic features (age, sex, household income, and marital status), one psychological feature (health consciousness), and six product features (price, easy operation, brand name, high performance, word-of-mouth communication, and cooperation with smartphones) constitute the network.

Findings from Probabilistic Inference. Based on the obtained Bayesian network structure, probabilistic inference was performed. Consequently, we arrived at consumer behavior knowledge regarding a blood-pressure monitor and a sleep monitor, as shown in Fig. 6. Consumers would maximally want to use a blood-pressure monitor, when the following product features were observed: high performance, reasonable price, easy operation, and brand name. These features increase the probability of the positive intention to use a blood-pressure monitor, from 26.7 % to 46.8 %. The same phenomena increase the probability of the positive intention to use a sleep monitor, from 11.5 % to 25.4 %. In short, these product features would positively influence the intent to purchase both a blood-pressure monitor and a sleep monitor. On the other hand, when people in their 50 s or older are observed (observation 2 in Fig. 6), the probability

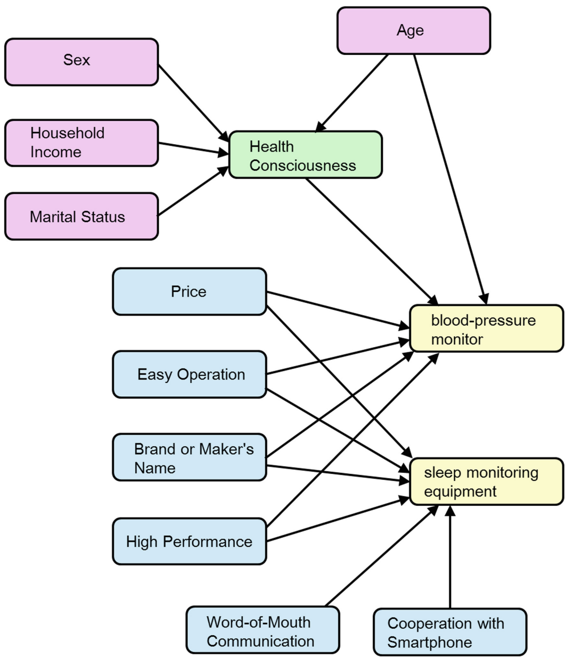


Fig. 5. A best Bayesian network structure obtained

of the purchase intention of a blood-pressure monitor would be positively increased to 39.4 %, but that of a sleep monitor would not change. In an opposite manner, when people learn about a product by word-of-mouth communication with friends or family members and the product cooperates well with their smartphones (observation 3 in Fig. 6), the probability of the intent to purchase a sleep monitor would be positively increased to 37.5 %, but that of a blood-pressure monitor would not change. Those differences derived from the asymmetry of the network structure shown in Fig. 5.

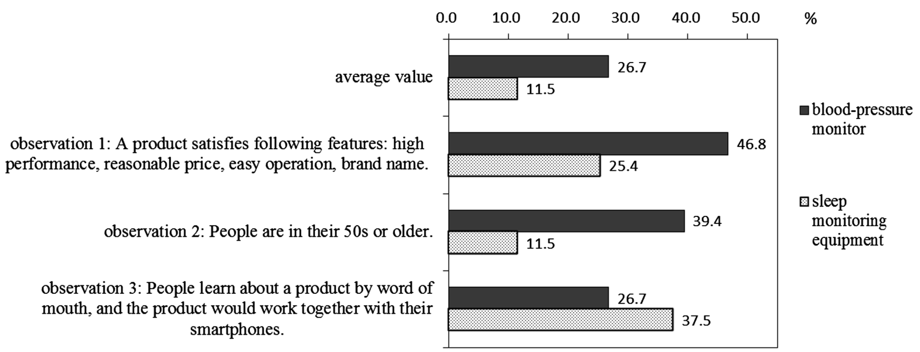


Fig. 6. Consumers' behavioral knowledge extracted by probabilistic inference

4.4 Text Mining

Subsequently, text mining was applied to answers to an open-ended question about desirable health-related home electronics. The number of respondents who answered at least a meaningful word in terms of content to the open-ended question was 1955, which is only 17.2 % of all respondents. We focused on the words "blood pressure" and "sleep," in order to integrate the results of text mining and the results of the Bayesian network analysis. As a result of the frequency analysis of nouns, "blood pressure" was ranked twelfth, while the value of *TF* (term frequency) was 48. Analogously, "sleep" was ranked twenty-fifth, while the *TF* value was 31.

Figure 7(A) and (B) indicate the results of the dependency parsing: (A) "blood pressure" was used as a keyword to catch a dependent word, and (B) "sleep" was also used to catch a dependent word. According to Fig. 7, the dependent words upon both keywords are similar to each other, including "record", "measure", and "control." However, in the case of "blood pressure", the most frequent dependent word is "measure," by a wide margin. Moreover, "blood pressure" has a small repertoire of dependent words, compared to "sleep." The healthcare electronics regarding "sleep" are firstly expected to be a device for measuring, recording, or controlling the quality and quantity of sleep. However, they are secondly expected to be equipment for achieving deep sleep, finding a sleep-related disease such as sleep apnea syndrome, or

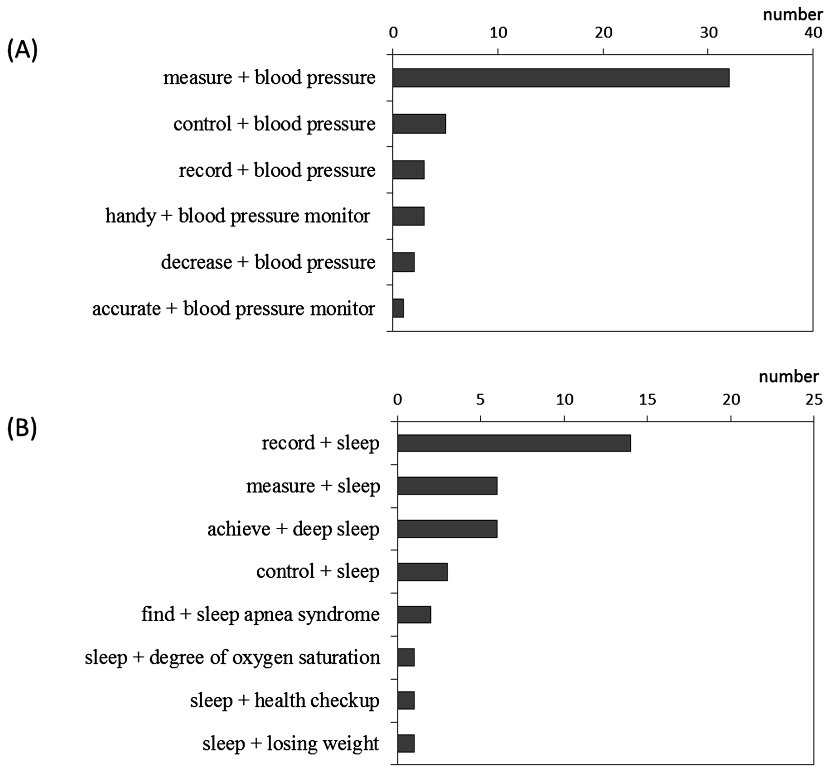


Fig. 7. Results of dependency parsing in text mining, on (A) blood pressure and (B) sleep

changing one's sleep pattern to lose weight. In terms of societal pervasiveness, a blood pressure monitor is a type of health equipment that has already spread among people, as shown in Fig. 2. On the other hand, the sleep monitor can be viewed a future product. According to the results of detailed text mining, people have begun to look for new routes of innovation regarding sleep monitors, riding on the strength of the diffusion of MHS.

4.5 Discussion

We have found several useful results for finding product innovation strategies based on marketing data. We concretely obtained consumer behavior knowledge related to the purchase intention of health-related home electronics. Using two different products as examples, a blood-pressure monitor and a sleep monitor, we finally obtained different behavior knowledge suitable for the nature of each product. In the case of a blood-pressure monitor, the desirable developmental direction is a device characterized by very accurate measurement or a handy and convenient measurement. In contrast, a sleep monitor has more choices for developmental strategies, not only focusing on accurate measurement, but also pursuing other potentials for achieving deep sleep, illuminating sleep-related abnormalities, or changing one's sleep pattern to lose weight.

These results show that our proposed method is effective in extracting consumer behavior knowledge from questionnaire data. Originally, Bayesian network modeling was one of the best methods of extracting knowledge from questionnaire data, since this modeling method is capable of handling nonlinear covariation between features and requires no designation of the objective variable and the explanatory variables in advance. However, this study demonstrated that appropriate feature selection is indispensable before starting to learn a Bayesian network structure for extraction of useful knowledge. Without proper feature selection, the knowledge obtained will be insufficient. This occurs when many attributes are highly associated with one another in a dataset, and some seem to form a cluster. Consequently, an excessive number of edges are inferred within such a cluster. This hampers the appropriate structure inference, since the inference (or learning) outcome is evaluated by using information criteria that restrict the number of edges, and many redundant but strong connections impede the weak yet appropriate connections. Therefore, proper feature selection is needed.

The proposed method based on Cramer's V values showed quite good performance. Since Cramer's V is a popular measure of association between two nominal variables, it is appropriate when analyzing categorical data in the marketing field.

5 Conclusions

In this study, we utilized multiple analyses including cluster analysis, graphical modeling, and text-mining, in order to extract real, useful consumer demands from survey data for the innovation of health-related home electronics. The effectiveness of the proposed method was verified by a case study using practical questionnaire data, while selecting a blood-pressure monitor and a sleep monitor as focused samples. The major originality of this research is attributed to the graphical modeling using Cramer's V feature selection and Bayesian network, which had been proposed by the author. This method enables us to solve the feature selection problem existing in survey data and to deal with multiple objective variables in one model. Moreover, it was indicated that integrating the results obtained by graphical modeling and the results obtained by text mining can reveal knowledge hidden in consumer awareness and behaviors. The results and the proposed method are certain to contribute to build up an innovative and favorable IoT-based healthcare system, from the product device side.

References

1. http://www.omron.com/
2. Ishino, Y.: Knowledge extraction of consumers' attitude and behavior: a case study of private medical insurance policy in Japan. Int. J. Web Eng. Technol. (in press)
3. Everitt, B.S.: Cluster Analysis, 3rd edn. Edward Arnold, London (1993)
4. Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: a review. ACM Comput. Surv. **31**(3), 264–323 (1999)
5. Heckerman, D., Geiger, D., Chickering, D.: Learning Bayesian networks: the combination of knowledge and statistical data. Mach. Learn. **20**, 197–243 (1995)
6. Friedman, N., Geiger, D., Goldszmidt, M.: Bayesian network classifiers. In: Provan, G., Langley, P., Smyth, P. (eds.) Machine Learning. Kluwer, Netherlands (1997)
7. Friedman, N., Nachman, I., Peér, D.: Learning Bayesian network structure from massive datasets: the sparse candidate algorithm. In: Proceedings 15th Conference on Uncertainty in Artificial Intelligence (UAI) (1999)
8. Jensen, F.V.: Bayesian Networks and Decision Graphs. Springer, New York (2001)
9. Jordan, M.I.: Graphical models. Stat. Sci. (Special Issue on Bayesian Statistics) **19**, 140–155 (2004)
10. Lauritzen, S.L.: Graphical Models. Clarendon Press, Oxford (1996)
11. Holmes, D.E., Jain, L.C. (eds.): Innovations in Bayesian Networks. SCI, vol. 156. Springer, Heidelberg (2008)
12. Chen, X., Anantha, G., Lin, X.: Improving Bayesian network structure learning with mutual information-based node ordering in the K2 algorithm. IEEE Trans. Knowl. Data Eng. **20**, 628–640 (2008)
13. Weiss, Y.: Correctness of local probability propagation in graphical models with loops. Neural Comput. **12**, 1–41 (2000)
14. Mooij, J., Kappen, H.: Sufficient conditions for convergence of the sum-product algorithm. IEEE Trans. Inf. Theory **53**, 4422–4437 (2007)
15. Blodgett, J.G., Anderson, R.D.: A Bayesian network model of the consumer complaint process. J. Serv. Res. **2**, 321–338 (2000)
16. Sheskin, D.J.: Handbook of Parametric and Nonparametric Statistical Procedures. CRC Press, Boca Raton (1997)
17. Manning, C.D., Schütze, H.: Foundations of Statistical Natural Language Processing. MIT Press, London (2002)

Knowledge Management in Healthcare

Ontology Driven Personal Health Knowledge Discovery

Hong Qing Yu(✉), Xia Zhao, Zhikun Deng, and Feng Dong

Department of Computer Science, University of Bedfordshir, Luton, UK
{hongqing.yu, xia.zhao, zhikun.deng, feng.dong}@beds.ac.uk

Abstract. With fast development of smart sensor devices and mobile applications, all different kinds of information related to humans can be founded on the Internet that can be seen as a universal data repository or called Web of Data. Health or healthcare related data are not exceptional in the Web of Data age. The most important and valuable data comes from IoT such as sensors and mobile activity tracking applications to support developing self-health risk detection and management applications. This paper presents a comprehensive ontology driven knowledge discovery framework in personal health domain, which aims to reason and discover health knowledge from various data sources of IoT. The framework contains a sensor oriented Personal Wellness Knowledge Ontology and data integration architecture to complete a whole lifecycle of health knowledge detecting and reasoning path. In addition, a cloud computing based parallel semantic lifting algorithm is described for illustrating the semantic data generation process in detail.

Keywords: Healthcare · Iot · Knowledge discovery · Semantic web and data mining

1 Introduction

This paper presents an innovated framework for managing and discovering people's health knowledge, which will exert influence on the future direction of people's self-care empowerment. With ageing population grows fast, healthcare and associated social welfare costs are growing exponentially and they will soon become unsustainable unless we change the way in which people are supported. In many cases, there is a need to shift medical care from institutions to home environment, and to transform healthcare from a system that is largely reactive – responding mainly when a person is sick – to the one that is much more proactive to supporting patients in self-management [1]. Recent research evidences show that chronic related patients who are more actively involved in their own healthcare receive better health outcomes [2]. Evidences also suggest that self-management skills can be developed and strengthened, even among those who are initially less confident, less motivated or have low levels of health literacy [3]. According to the analysis from our current experience in MyHealthAvatar, one of the key factors of success in patient empowerment is to allow them to gain valuable knowledge from their own data, bringing them tangible benefits.

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 649–663, 2015.
DOI: 10.1007/978-3-319-21009-4_48

Our solution is an attempt to maximise the benefit of the data that are made available by the fast growing development on Internet of Thing (IoT), especially the self-monitoring sensors aiming at exploring this newly arising opportunity for healthcare. There has been indeed growing interest in the "initiative of self-monitoring", evidenced by the sharp market expansion in life-logging devices and apps. These sensors are capable of constantly monitoring personal health behaviours and activities (e.g. walking, calories, heart rate, and diet), leading to unprecedented opportunities in self-care. Correspondingly, significant research efforts have started to harvest and integrate the sensors for long-term health data collections – examples include MyHealthAvatar [4] and MyLifeHub [5]. Such a long-term data collection is extremely valuable to individualised disease prediction and prevention, and to promoting healthy lifestyles.

However, this "self-monitoring initiative" demands significant motivation and commitments from the users, as they can only achieve meaningful size of data collection by constantly wearing and maintaining the sensors on a daily basis for a long-term duration. Hence it is vital that we encourage the user commitments by feeding back to them useful information extracted from the data that they have self-collected, creating incentives that can further inspire them to continue the self-monitoring. The highly heterogeneous and dynamic nature of the data brings new challenges to the technologies. In order to resolve the challenges, the first step is to enable meaningfully storing and querying the data for having basic data analysis and reasoning. Since the highly heterogeneous nature of IoT, we need to deal with complex queries than traditional single data resource scenarios. Take an example to illustrate the complex query required by health self-management application.

An application user uses advanced sensors for dynamically recording his activities and physiological indicators. These records may include day-based steps, sleep qualities, weight, exercises (running, walking or cycling) and blood pressure. Each of these record data can be retrieved through individual queries. In addition, user static profile data are registered initially such as date of birth, habit (smoke and drink) and height. However, individual value cannot represent the health situation of the user. The health situation requires to be indicated by having complex queries. There are two types of complexity:

Multi-combination query: For example, the BMI status needs to know the weight and height from two different sources of dynamic sensor and static profile. More examples such as risk examine functions provided by Framingham Heart Study[1] require even more complex queries that need to combine more than two data resources.

Reasoning-based query: The query results can identify links between activities and health status. For example, to check which activities triggered the consequence of user's hypertension condition? The reasoning method in this case may compose some complex queries together such as time considered sleep query, exercise query and steps query.

[1] https://www.framinghamheartstudy.org/.

2 Research Background and Related Work

2.1 Background

IoT data can be collected through sensor or application data APIs or directly embedded in the healthcare system. Specially, in the last decade, Body Sensor Networks [6], or BSNs are developed to remotely collect data and upload vital statistics to servers over the Internet majorly because the high demands for efficient health monitoring, which forced the health and wellness industry to embrace modern technological advances [7]. BSNs can efficiently provide monitored and recorded data, when communicate to the suitable systems.

Each individual data resource can only presents partial aspect of the health information. For example, move related sensors can normally monitor human movement and possible calories consumption but not social activates, weight or mood you currently are, which may be in your social media communications or definitely required human interactions. Therefore, the integration of the heterogeneous data is the foundation to support more advanced knowledge discovery and reasoning.

Luckily, the health related data integration work has been focused by some research projects in last couple of years. For example, MyHealthAvatar framework is a proof of concept EU funded 3 million euros project for providing a digital representation of patient health status. It is designed as a lifetime companion for individual citizens that facilitates the collection of, and access to, long-term health-status information that includes citizens' social and sensor data together with major data resources from traditional healthcare organisations. Currently, the integrated are stored in a cloud and NoSQL based data repository with Web APIs to query and access. This paper will focus on the semantic reasoning framework that will extract health event knowledge from the IoT raw data into semantic representation and to be queried to perform further semantic based reasoning tasks.

Taking diabetes condition as example, it is a major cause of cardiovascular disease; the most common reason for commencing renal dialysis; the most common cause of blindness in people under the age of 60 and non-traumatic lower-limb amputation. Therefore the goal of the risk detection is to enable:

(1) To summarise the physiological indicator data to gain the knowledge of risk level and trend of "John" on diabetes condition.
(2) To semantically represent knowledge that will be meaningfully and efficiently to support reasoning on risk levels of other related diseases, such as cardiovascular, commencing renal dialysis and blindness.
(3) To understand the connections between lifestyle and condition risks in order providing an incentive lifestyle facilities for "John" helping on health self-managements.

2.2 Related Work

Ontologies for describing personal health conditions have been proposed recently to utilize the advanced semantic capability helping on the personalised healthcare management and self-empowerment. The early stage, the ontology definitions tend to represent existing personal health information (e.g. PHR) at medical organistions e.g. hospitals with semantic indications. For example, [8] describes a ontology model that extends original work from [9] achieving semantic interoperability by developing a specific ontology for active PHRs with syntactic interoperability using Health Level 7 (HL7) standards for data and document exchange. Reference [10] proposes a Case Profile Ontology (CPO), which defined by OWL-DL compliant ontology and used in the K4CARE project[2] for providing a formal representation of a comprehensive chronic disease associated personalisation healthcare concept set at home such as syndromes, social issues, signs and symptoms, problem assessments and interventions and the relationships and constraints between these concepts. The data on the web oriented ontology presented in [11] that describes a Patient-based Web Written Message Ontology (PWMO) in order to analysis patient social behavior on the web for supporting patient condition conclusion. Reference [12] proposes a personal wellness information model (PWIM) that describes the domain of personal lifestyle and includes concepts of activities, behaviors and choices that affect the person's daily life. In the meantime, it aims to support personal health self-management concepts. PWIM focuses more on lifestyle management in contrast to CPO concentrating on patient information management. Most recently, [13] presents a H-event ontology for supporting discovery healthcare related event and treating the event as a gaining knowledge for facilitating personalised healthcare and decision making. In this paper we extends the H-event ontology by specifying the sensor information to detect significant event fact. We also detailed describes the mechanism and algorithms that designed for implementing the overall platform that works in MyHealthAvatar project. Event-based knowledge extraction is the most efficient and understandable methodology to get meaningful information from unstructured data [16]. An important feature of an event is that it can specify a condition that could be measured or detected by a set of unique attributes that are differentiated from other phenomenal. In recently years, many event-based knowledge extraction methods have been developed in different areas such as arts [15], economic [17], social networks [18], and biomedical research [19]. Most of these processes are supported by semantic technologies.

2.3 Research Challenges

(1) Semantic modelling of personalised health knowledge to support further analysis and reasoning.

To have a huge size of machine-not-understandable data is not ideal for data analytics process. Especially, the meaning and relations of the data will be vital to decrease knowledge discovery workloads. Semantic technology is the useful tool to popular the semantic value to abstract the original data for making a more meaningful data structure.

Although convergence of large data and semantic technology is not straight forward because the semantic layer of LOD (Linked Open Data) and the large-scale data storage

[2] http://www.k4care.net.

do not get along easily, semantic standards still support a huge advantage to have a machine-readable semantic layer over structured as well as unstructured data, which hence automate data analysis tasks. Meanwhile, the sheer data size envisioned by large data denies certain computationally expensive semantic technologies, rendering the latter much less efficient than their performance on relatively small datasets. Therefore, the trade-off balance between which data are required to represent with semantics and which one are not is an important factor to implement the data extraction process [20].

Therefore, well-structured semantic model should be defined at first place.

(2) The knowledge detecting mechanism to observe risk factors from a big and day-to-day data repository.

The unique feature of acquiring healthcare knowledge is that some of them cannot be derived directly from single data record or even a small size of data collection. For example, to understand one particular person's blood sugar level is not as direct as just evaluating one single test result because blood sugar level is very variable that depends on person's changes on diary and activities. However, evaluating statistics of a long period test records can recognize the high blood sugar level fact. The same scenario also can be applied to check side effects of medications. The most challenge task to meet the long period data analysis is that all different knowledge facts require to be assessed by different evaluation methods. Therefore, well-defined knowledge extraction methodologies are required to enable dealing with different evaluation requirements at run-time.

(3) Dynamically conducting reasoning rules from both semantic data and raw data.

The unique feature of acquiring healthcare knowledge is that some of them cannot be derived directly from single data record or even a small size of data collection. For example, to understand whether a particular activities or food taken associated to one or more particular symptoms. The same scenario also can be applied to check side effects of medications. However, the most challenge task to meet the long period data analysis is that all different knowledge facts require to be assessed by different mining methods. Therefore, well-defined knowledge reasoning methodologies are required to enable dealing with different evaluation requirements at run-time.

By considering the above challenges, we investigated an Ontology Driven Personal Health Knowledge Discovery process for solve them.

3 Event-Based Personal Wellness Knowledge Ontology (WKO)

The Event-based Personal Wellness Knowledge Ontology (WKO) is the extension version of H-event ontology introduced by [13]. Originally, the H-event ontology focuses on health condition related concepts and their linked relations without specifying the way to model daily actions and risk analysis capabilities. WKO majorly extends the H-event ontology in two areas: (1) detailed semantic concepts to model daily actions of each individual that are monitored by sensor devices and (2) risk level assessment concepts to support risk analysis features. The specifications of WKO semantic model are described as following:

- **Event** is the concept to identify an unusual situation occurred to the user, the event is the supper class of (appearing a) **Symptom**, (taking a) **Treatment**, (diagnosed a) **Health Condition** and (Having a) **Daily Action**, which extends the existing event ontology[3]. Each event associates to a particular time point on user's MyHealthAvatar Timeline. In addition, Event is the central point of the whole ontology, which can be detected from the large dataset collected from the system pre-processing.
- **Person** is the concept to describe a MyHealthAvatar user using FOAF ontology[4]. The FOAF ontology includes all possible aspects about a general profile of a person such as name, gender and contact information.
- **Daily Action** is a subclass of **Event** concept to identify the daily activities or health related facts that could affect or useful to understand the user's health situation that showed on the timeline. The activity and fact type can be identified by one or more specific sensors. For example the activity type can include "Running", "Driving", and "Shopping". The fact type can be running distance, calories consumptions and steps numbers.
  - **Sensor** specifies a certain device or application that could be used for collecting health related data or activity data.
  - **(Significant) Activity** specifies an action that detected by one or more sensor devices/applications may have lots of different types such as exercise, travelling, working or drinking. The significant activity is a core concept of the ontology for reasoning of health knowledge.
  - **Fact** describes useful detail information of activity such calories consumption, distances and frequency.
- **My Timeline** is the concept to define a time (4th) dimension data representation of a MyHealthAvatar user to specify the exact time point of each event, which is modified from time ontology[5].
- **Place** is the concept define in place ontology[6] to identify accurate location of a detected healthcare event, such as hospital, home or a particular place.
- **Annotation** is a semantic vocabulary link of an event. The annotations should use controlled vocabularies or semantic identifiers to define the meaning of events in order to enable machine to understand for enhancing knowledge discovery.
- **Symptom** is a subclass of Event concept to present the unusual health related affects that are detected and concluded from the user's data. As same as all other events, the symptoms have to have a time stamp on user's timeline.
- **Treatment** is a subclass of Event concept for recording the treatments that have been taken by the user from medical health organization or user-self. The treatment refers to any medical actions that have been done to the user such as taking dugs,

[3] http://motools.sourceforge.net/event/event.html.

[4] http://xmlns.com/foaf/spec/.

[5] http://www.w3.org/TR/owl-time/.

[6] http://vocab.org/places/schema.html.

operation and physical and mind therapies. In addition, the treatment requires identifying the exact time point on user's timeline.

- **Health Condition** is a subclass of Event concept and presents the medical situations that were diagnosed in pass according the user's timeline or will be a potential risk for the user. The Health Condition, Treatment and Symptom concepts structure a triangle relation that could be a very valuable knowledge for individual user or a group of users.

  - **Risk** is a concept to evaluate the possibility or progress levels to a particular health condition.
  - **Level** is a concept to indicate the risk volume for progressing to a particular health condition. Scale of the level is [0, 1, 2, 3, 4, 5]. High value indicates high-level risk (Fig. 1).

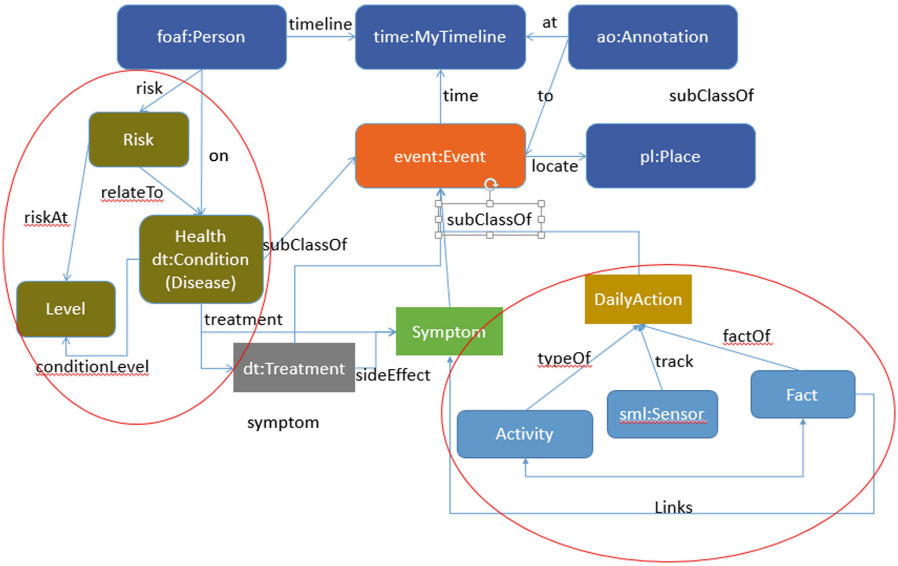


Fig. 1. WKO Ontology top level concepts

4 Event Based Knowledge Discovery and Semantic Lifting Process

4.1 Overall Architecture

The whole event based knowledge semantic lifting process includes 2 major components (see Fig. 2)

- **Data Aggregator** harvests the heterogeneous health related data from different resources in to an aggregated data repository. The data aggregator is composed by two sub-components namely **Data importer** and **NoSQL data repository.**

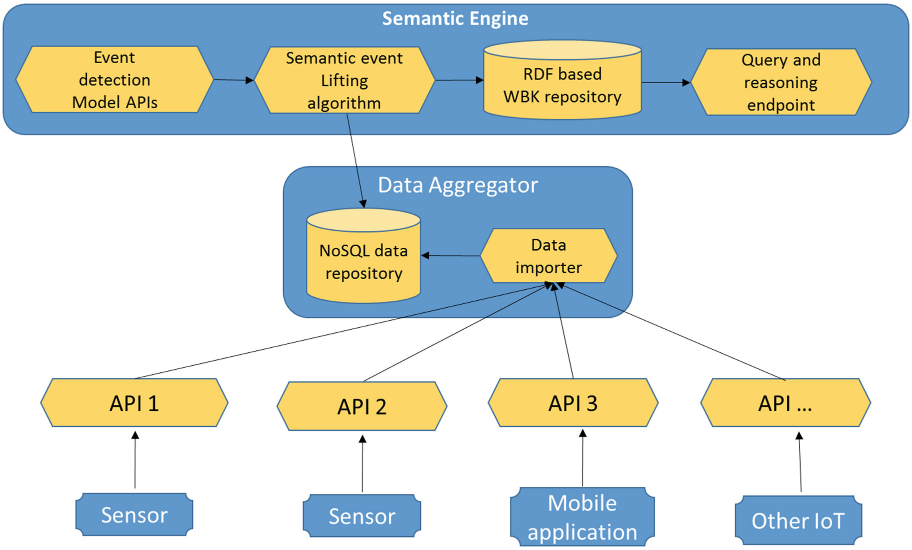


Fig. 2. The overall architecture of event detection and semantic lifting process

- **Data importer** pulls the data from individual sensor devices from their provided Web APIs. The data importer is the Data Processing Engine that can crawl, pre-process and integrate heterogeneous data from Web APIs.
- **NoSQL repository** takes charge of maintaining the big datasets that are harvested from various data resources pushed by **Data importer** and also including data generated from client application. At the moment, the supported NoSQL repository configuration in the architecture is Cassandra. We consider the following aspects when choosing between SQL and NoSQL databases for archiving our goal. (1) Natural of the data to be stored: To store and access the lifelong information of each citizen in MyHealthAvatar, there will be large volume of data in unstructured, semi-structured and structured formats. For example, data collected from wearable devices and social networks may come with different formats (e.g. XML and JSON) and involves multiple data types (e.g. objects, lists and customised data types), which are difficult to be incorporated into the rows/columns table style in SQL. Furthermore, data from different resources with unknown structure are likely to be added in future. It is not feasible to define the data type and structured beforehand as required by SQL databases. (2) Data analytics: It is important to perform data analytics and possibly near real-time analytics in MyHealthAvatar in order to support event detection. (3) System scalability and performance: we foresee the future of such systems will manage large amount of user data with high demands on updating and reading. It is vital the system can scale and adapt the distributed computation paradigms (e.g. cloud computing) when the volume of the data grows. (4) Agility of the application development: MyHealthAvatar aims to collect and access the lifelong health data for citizens. The uncertainty of the data types

from different source centers requires that the development is done in iterations and be flexible and dynamic when new data resources are added. The designation of the NoSQL data schema is represented in Fig. 3. The data storing mechanism is user centric and each column of the value (Knowledge Base) will linked to an entity or target in a timeline-based id.

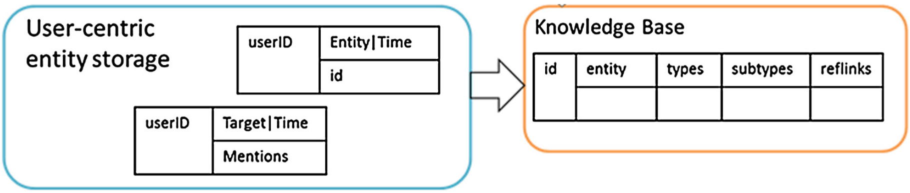


Fig. 3. NoSQL data persistency schema

- **Semantic Engine** provides the functions to lifting the health knowledge into the semantic data repository and to be queried and reasoned by client applications through semantic accessing endpoints. The semantic engine includes 4 major components.
  - **Semantic event lifting algorithm** is the core of the process for controlling the detection of state changing by applying a specified data analysis algorithm. A specified data analysis algorithm may be applied to one or more suitable data pool, but most analysis algorithms are designed for a particular data resource such as step analysis algorithm that illustrated in Algorithm 1. Once the analysis algorithm detected the state changes, these changes are seen as events and semantically lifted to the semantic data repository. The semantic lifting algorithm is described in Algorithm 1.
  - **Semantic data repository** is implemented by OWL-based RDF repository (OWLIM). The repository stores the RDF triples following the Wellness Knowledge Ontology (WKO) that is defined in Sect. 3. One of the major feature of OWLIM repository supports context closure that can group individual user's triples by individual user's id for enabling efficiently querying different user's data.
  - **Query and reasoning endpoint** provides two basic functions: (1) accepting SPARQL query for normal OWL-based RDF responses and (2) accepting SPARQL-based XML query template for performing complex query and reasoning.
  - **Event detection model APIs** support computation models that are used for detecting the event from the NoSQL repository. For example to identify a location changes, abnormal activates or physiological indicators.

4.2 Semantic Event Lifting Algorithm

Before starting introduce the algorithm, the event definition is defined as following:

An event is a tuple $<\mathbf{ID}, \mathbf{Dsc}, \mathbf{CM}, \mathbf{T}, \mathbf{S}>$

Where:

ID is the identifier of the event which related to hearth knowledge.

Dsc is the description of the purpose of the event e.g. detecting the last month average blood pressure value or frequency of high blood pressure phenomenon in past month.

CM is an URI that accesses to an event discovery computation model that is provided as an API in the system implemented or defined by medical professionals. The API has input parameters and output schema. For instance, calculating frequency of high blood pressure in past month, the inputs will be the user id and blood pressure values in last 30 days; the outputs will be a fuzzy value in a range of defined terms < none, low, normal, high, risk > .

T specifies the output of the **CM** should link to the semantic terminology defined in the WKO, e.g. the output of frequency of high blood pressure should be lifted as terms of fact and symptom in the WKO specification. Then the semantic lifting algorithm can correctly lift the result into semantic repository as knowledge rather than data information.

S specifies the sensors or applications that could be used for retrieving the required data by using the sensor Id number.

The event will also registered in the semantic repository. Below is an example of event Abox instances (id is a URI under MHA domain) registration to the system.

<mha:event0001> <hasDsc> "calculating frequency of high blood pressure in past month"

<mha:event0001> <hasCM> <http://example.bloodpressure/calculation>

<mha:event0001> <hasT> <WKO:fact>

<mha:event0001> <hasT> <WKO:sympotm>

<mha:event0001> <hasS> <mha/user0001/sensor:0001>

<http://example.bloodpressure/calculation> <hasInput> <http://dbpedia.org/resource/Blood_pressure>

<http://example.bloodpressure/calculation> <hasInput> <http://dbpedia.org/resource/Blood_pressure>

<http://dbpedia.org/resource/Blood_pressure> <hasPeriod> "30"
…

Based on the event definition, a cloud parallel computing based semantic event lifting algorithm is developed as Algorithm 1:

Algorithm 1: semantic event lifting for a single user

This algorithm will apply all registered event computation models to discovery health knowledge about a user and lift the outcomes to the semantic representations according to WKBO specification.

```
Input: E[]=Array<EventURI>, UserURI
Output: RDF triples
```

```
Parallel each event E[i] to a cloud computation node

    E_graph=SPARQLQueryTuple(E.i)

    Input[]=E_graph.hasInput.term

    int Required_period = E_graph.hasInput.period

    CM_URI=E.graph.hasCM

    Input_value[] = Array<Vector>

    Invoke_URI
    Output_value[]

    From j=0 to input.length()
        Input_value[j]=NoSQLValueQuery(Input.j, Requried_period)
        Invoke_URI=InvokeURIGeneration(Input_value[j], CM_URI)

    Output[]=Invocation(Invoke_URI)

    Return RDFTripleLift(Output[], E_graph.term)
```

5 Discussions

At current stage, we developed the very basic infrastructure and functions try to build a foundation to a comprehensive solution for health knowledge discovery. On technical side, there are still some open research challenges. One of the key challenge will come from data reliability and usability – while the self-monitoring sensors do provide valuable information about personal activities and lifestyles, they are normally low-cost and tend to have noisy and incomplete data.

Concerns have been raised over their reliability and accountability. In the future, the semantic engine will be extended to investigate novel methods to tackle data reliability and usability issues by looking into new personalised techniques based on the current state of the art to handle noisy and missing data, taking into account user profiles that will be made available through the data from the MyHealthAvatar platform. One possible direction could be based on uncertain data analysis, uncertain data clustering and classification, dynamic Bayesian network, which will be designed to accommodate low data quality, making use of the latest evaluation outcomes over self-monitoring sensors in MyHealthAvatar priors for the data mining. Notably, particular attentions will be paid towards new methods that are able to progressively build individualised priors over the historical knowledge stored at WKBO and make use of these priors to consolidate WKPO through incremental evolutions. We shall also consider new methods that can integrate user feedback into the WKBO evolution by applying penalties to the computing outcomes that are not agreed by the users.

Noisy data handling and data imputation have been a hot research topic especially in light of the significant increase of low-cost sensors. There is a set of methods can be adopted to do imputation task such as Prediction Mean Matching Imputation, KNN and Regression methods [21] and Attribute Selection, Smart Tokens and Probabilistic Noisy Identification methods for removal noisy data [22, 23]. Also, there exist a spectrum of methods for representation and manipulation of uncertainty, such as probability density function, fuzzy sets, belief function, or interval sets, uncertainty propagation and sensitivity analysis, etc. [24]. Meanwhile, techniques have also been experimented to progressively update data mining models upon the arrival of new data, e.g. online machine learning techniques [25]. Also, many personal recommendation systems have taken user profiles and feedback into consideration in their data mining processes [26, 27].

In brief, while many relevant technologies do exist, they have yet been fully investigated, evaluated and adapted under the scenarios of self-monitoring for healthcare. With knowledge of personal health profiles through a long term self-monitoring, our research work will lead to potentials in a wide range of healthcare applications. In addition, it will be an experimental study for big data research by combining data mining and semantic technologies for healthcare, offering guidelines and references for future research in similar directions.

6 Conclusions and Future Work

In this paper, we introduced a Ontology based framework for health related knowledge discovery through semantic abstractions of integrated data resources from IoT. The framework works with a hybrid data repository combining NoSQL and RDF storage. The proposed approach mainly contribute to solve the knowledge discovery required complex data query and reasoning by tackling three major challenges that are raised in healthcare data analysis domain namely: (1) heterogeneous data integration, (2) extraction knowledge by using long period monitoring data and (3) dynamically conducting reasoning rules from both semantic data and raw data. The whole approach relies on two major processes of entity extraction and ontology based detection.

The first data integration process imports and extracts IoT raw data from various datasets that are retrieved through the Web or Web APIs. These data majorly include sensor data and mobile applications.

The second semantic lifting process further extracts explicit healthcare knowledge about the user based on the extension of H-event ontology called WKO which enables detecting specific activity oriented healthcare related events that support monitoring and analysis features to MyHealthAvator platform. The detected knowledge finally builds its own knowledge bases for individual user as well as group users for healthcare purposes. Extracted data in two different levels are stored in hybrid data repository using NoSQL and RDF triple storages for supporting different levels of data query and analytics requests. Currently, the approach works as backend bone to the MyHealthAvatar platform.

In the future, our work will focus on two research directions. The first one is to deal with raw data reliability such as data inaccuracy and missing. The second one is to evaluate and improve our cloud based algorithm for searching a long-term solution for enhancing the scalability of the approach.

Acknowledgments. This work is supported in part by the European Commission under Grant FP7-ICT-9-5.2-VPH-600929 within the MyHealthAvatar project: A Demonstration of 4D Digital Avatar Infrastructure for Access of Complete Patient Information.

References

1. Wagner, E.H.: Chronic disease management: what will it take to improve care for chronic illness? Effective Clin. Pract. **1**(1), 2–4 (1998)
2. Greene, J., Hibbard, J.H.: Why does patient activation matter? An examination of the relationships between patient activation and health-related outcomes. J. Gen. Intern. Med. **27** (5), 520 (2012)
3. Hibbard, J.H., Greene, J.: What the evidence shows about patient activation: better health outcomes and care experiences; fewer data on costs. Health Aff. (Millwood) **32**(2), 207–214 (2013)
4. MyHealthAvatar. http://myhealthavatar.eu
5. MyLifeHub. http://gow.epsrc.ac.uk/NGBOViewGrant.aspx?GrantRef=EP/L023830/1
6. Garg, M.K., Kim, D., Turaga, D.S., Prabhakaran, B.: Multimodal analysis of body sensor network data streams for real-time healthcare. In: Proceedings of the International Conference on Multimedia Information Retrieval (MIR 2010), pp. 469–478. ACM, New York (2010)
7. Bauschlicher, D., Bauschlicher, S., ElAarag, H.: Framework for the integration of body sensor networks and social networks to improve health awareness. In: Proceedings of the 14th Communications and Networking Symposium (CNS 2011), pp. 19–26. Society for Computer Simulation International, San Diego (2011)
8. Plastiras, P., O'Sullivan, D., Weller, P.: An ontology-driven information model for interoperability of personal and electronic health records. In: Paper presented at the Proceedings of eTELEMED, the Sixth International Conference on eHealth, Telemedicine, and Social Medicine, Barcelona, 24–27 March 2014

9. Puustjärvi, J., Puustjärvi, L.: Designing and implementing an active personal health record system. In: The Third International Conference on e-Health, Telemedicine and Social medication, pp. 102–107 (2011)
10. Riaño, D., Real, F., López-Vallverdú, J.A., Campana, F., Ercolani, S., Mecocci, P., Annicchiarico, R., Caltagirone, C.: An ontology-based personalization of health-care knowledge to support clinical decisions for chronically ill patients. J. Biomed. Inform. **45**(3), 429–446 (2012)
11. Marina, S., David, S.: Building a patient-based ontology for user-written web messages. In: Proceedings of the International Conference Recent Advances in Natural Language Processing 2011, Association for Computational Linguistics, pp. 758–763 (2011)
12. Seppälä, A., Nykänen, P., Ruotsalainen, P.: Development of personal wellness information model for pervasive healthcare. J. Comput. Netw. Commun. 2012, 10 (2012). Article ID 596749
13. Yu, H.Q., Zhao, X., Zhen, X., Dong, F., Liu, E., Clapworthy, G.: Healthcare-Event driven semantic knowledge extraction with hybrid data repository. In: 2014 Fourth International Conference on Innovative Computing Technology (INTECH), pp. 13–18, 13–15 August 2014
14. Chen, H.L.: A socio-technology perspective of museum practitioners' image-using behaviors. Electron. Libr. **25**(1), 18–35 (2007)
15. Lin, C., Yen, C., Hong, J., Cruz-Lara, S.: Event-based knowledge extraction from free-text descriptions for art images by using semantic role labeling approaches. J. Electron. Libr. **26** (2), 215–225 (2009)
16. Alphonse, E., Aubin, S., Bisson, G., Hamon, T., Lagarrigue, R., Nazarenko, A., Manine, A., Nédellec, C., Ould, M., Vetah, A., Poibeau, T., Weissenbacher, D.: Event-based information extraction for the biomedical domain: the caderige project. In: Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and its Applications (2004)
17. Hogenboom, A., Hogenboom, F., Frasincar, F., Schouten, K., Meer, O.: Semantics-based information extraction for detecting economic events. Multimedia Tools Appl. **64**(1), 27–52 (2013)
18. Rattenbury, T., Good, N., Naaman, M.: Towards automatic extraction of event and place semantics from flickr tags. In: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2007), pp. 103–110. ACM, New York (2007)
19. Kokkinakis, D.: Initial experiments of medication event extraction using frame semantics. In: Scandinavian Conference on Health Informatics (SHI 2012), pp. 41–47 (2012)
20. Hu, B., Carvalho, N., Laera, L., Matsutsuka, T.: Towards big linked data: a large-scale, distributed semantic data storage. In: Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services (IIWAS 2012) (2012)
21. Suthar, B., Patel, H., Goswami, A.: A survey: classification of imputation methods in data mining. Int. J. Emerg. Technol. Adv. Eng. ISSN **2**(1), 2250–2459 (2012). ISSN 2250-2459
22. Tamilselvi, J.J., Saravanan, V.: Handling noisy data using attribute selection and smart tokens. In: International Conference on Computer Science and Information Technology, pp. 770–774, 29 August–2 September 2008
23. Kubica, J., Moore, A.: Probabilistic noise identification and data cleaning. In: Third IEEE International Conference on Data Mining, pp. 131–138, 19–22 November 2003
24. Frey, H., Patil, S.: Identification and review of sensitivity analysis methods. Risk Anal. **22** (3), 553–578 (2002)
25. Blum, A.: On-line algorithms in machine learning. Technical Report CMU-CS-97, Carnegie Mellon University (1997)

26. Lops, P., de Gemmis, M., Semeraro, G.: Content-based recommender systems: state of the art and trends. In: Ricci, F., et al. (eds.) Recommender Systems Handbook, pp. 73–105. Springer Science and Business Media, LLC, New York (2011). doi:10.1007/978-0-387-85820-3_3
27. Sun, X., Khoshgoftaar, T.M.: A survey of collaborative filtering techniques. Adv. Artif. Intell. 2009, 19 (2009). Article ID 421425

Design of Query-Driven System for Time-Utility Based Data Mining on Medical Data

G. Arumugam[1] and V.K. Vijayakumar[2(✉)]

[1] Department of Computer Science, Madurai Kamaraj University, Madurai, Tamil Nadu, India
gurusamyarumugam@gmail.com

[2] Department of Computer Science, Sourashtra College, Madurai, Tamil Nadu, India
vijayakumar.sou.college@gmail.com

Abstract. Association rule mining(ARM) techniques search for groups of frequently co-occurring items (i.e., frequent itemset) in a market-basket transaction database and convert these groups into business-oriented rules. The problem of ARM will gain momentum when it is attached with the time of transaction. High utility itemset mining is a research area of utility based data mining, aimed at finding itemsets that contribute most to the total utility. The association of time and utility on frequent itemsets gives a novel approach to efficiently capture the transactions for getting better predictions and planning for an enterprise. Previous research has focused mainly on how to obtain exhaustive lists of association rules. However, users often prefer a quick response to targeted queries. To accelerate the processing of such queries, a query-driven system called TD-FVAUFM (Time-Dependent Fast Value Added Utility Frequent Mining) is proposed in this paper. It performs data preprocessing steps on the given database and the resultant database is converted in the form of an *itemset tree*, a compact data structure suitable for query response. The proposed system is applied on a medical database containing patient's records. It generates association rules that predict possible diseases with risk factor and frequency with respect to time. Experiments indicate that the targeted queries are answered in a time that is roughly linear in the number of transactions.

Keywords: Itemset tree · Time-Dependent utility frequent itemset · Utility based data mining · Association rules · Medical database

1 Introduction

Data mining is the process of knowledge discovery from patterns generated by applying various functionalities associated with it. The entire life cycle of knowledge discovery includes steps such as data cleaning, data integration, data reduction, data transformation, data mining, pattern evaluation, and knowledge presentation. Data mining is used to build predictive and descriptive models. A predictive model is used to

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 664–682, 2015.
DOI: 10.1007/978-3-319-21009-4_49

explicitly predict values. Descriptive models on the other hand describe patterns in the existing data. It characterizes the general properties of the data in the database [5].

Predictive data mining techniques are used in decision making where the results out of the decision give us better experience and wide knowledge. The experience and knowledge are used to update the predictions for the best results. Hence it is a knowledge learning cycle as shown in Fig. 1.

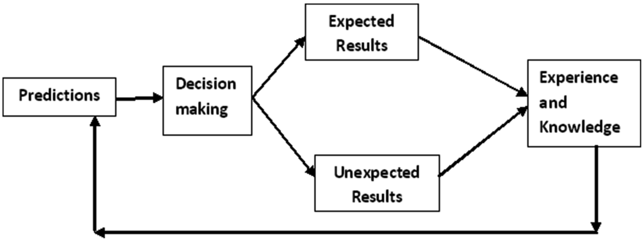


Fig. 1. Knowledge Learning Cycle

The success or failure of an event is associated with so many factors. In normal day-to-day life things are associated together which give us a rich knowledge and experience. It is a well known fact that for every action there is a proportionate consequence. In other words, all events in the real world can be converted in the form of a rule " if (antecedent) then (consequence)" (i.e.,) if (A is True) then (B is True). In other words an event can be represented through implications of the form A ⇒ B. Such implications are known as *association rules*. The validity of a rule is usually measured using two measures namely support and confidence. But in this paper, two more measures namely time and utility are introduced to validate such rules as shown in Fig. 2.

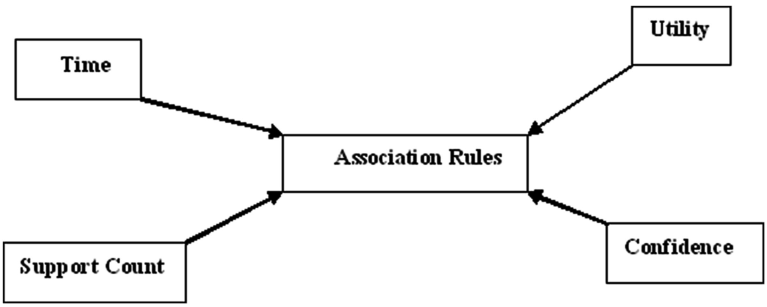


Fig. 2. Four dimensions of Association Rules

In this paper, a query driven system is proposed and the system is applied on a medical transactional database containing patient's records for disease diagnosis [2, 7, 10, 11]. The proposed system analyzes patient's records and generates association rules of the form {Symptoms} ⇒ {Disease}, with three measures viz., *confidence* (probability of a patient getting affected by a disease with the given symptoms), *support count* (how often the patient is affected by a disease with the given symptoms) and *utility* (The risk factor about the disease).

2 Background Study

2.1 Itemset Trees for Targeted Association Querying

The paper [6] suggests an algorithm "CONSTRUCT" for keeping the records of the database in the form of an itemset tree. Consider a database D = {{1,2,5}, {1,2,4}, {1,3},{1,2,3,5},{1,2,3}}. The algorithm builds the itemset tree incrementally, through a series of market basket insertions of the database D as shown in Fig. 3.

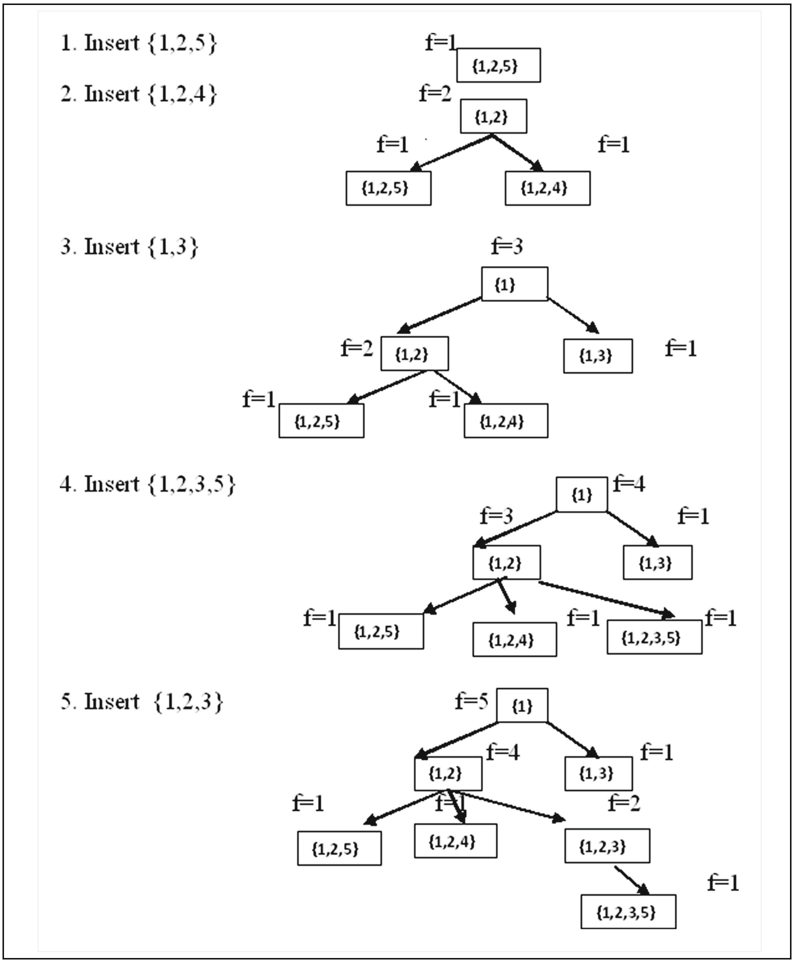


Fig. 3. Itemset tree construction using the "CONSTRUCT" algorithm

An ancestor of an itemset C can be viewed as an uninterrupted sequence of the smallest items of C. Here {1,2} is an *ancestor* of {1,2,5}. The *largest common ancestor* of {1,2,5} and {1,2,4} is {1,2}. Note that {1} is not the largest common ancestor of these two itemsets. Further {1,2,5} is a *child* of {1,2} but {1,2,3,5} is not a child of {1,2,5} because {1,2,5} is not an ancestor of {1,2,3,5}.

The *frequency* of a node in an itemset tree equals the number of market baskets that have passed through this node during the construction of the itemset tree. The frequency of the child is smaller than the frequency of the parent.

2.2 Utility Based Data Mining

Utility based data mining [8, 12–14, 16] gives a new dimension to frequent itemset mining problems. High-utility itemset mining model was defined by Yao, Hamilton and Butz [15]. The objective of high utility itemset mining is to find all itemsets that give utility greater than or equal to a user specified threshold. The utility values of various components are illustrated as shown below with the help of a small database given in Tables 1 and 2.

Table 1. External Utilities of Items in the Set I = {1,2,3,4,5}

Item	1	2	3	4	5
External Utility (or)Profit (y_k)	5	2	1	4	3

Table 2. Market Basket Database D, with 5 Transactions and 5 Distinct Items

TID	Items bought	Internal utility (or) quantity (x_k)
T_1	{1,2,5}	{2,1,4}
T_2	{1,2,4}	{5,3,6}
T_3	{1,3}	{4,3}
T_4	{1,2,3,5}	{1,1,1,1}
T_5	{1,2,3}	{2,2,2}

- The utility of item 5 in itemset S = {2,5} is calculated as u(5,{2,5}) = u(5,T1) + u(5, T4) = 4*3 + 1*3 = 15.
- The utility of itemset S = {1,5} in D is u({1,5}) = u({1,5},T1) + u({1,5},T4) = 30.
- The utility of transaction T5 is u(T5) = u({1},T5) + u({2},T5) + u({3},T5) = 16.
- The utility of database D is u(D) = u(T_1) + ... + u(T_5) = 24 + 55 + 23 + 11 + 16 = 129.
- The utility share of itemset {1,5} in database D is calculated as U({1,5}) = u({1,5}) / u(D) = 30/129 = 0.2325 = 23.25 % .

FUFM algorithm. The algorithm FUFM (Fast Utility-Frequent Mining) [9] finds utility-frequent itemsets using the existing frequent itemset mining algorithms [1, 3, 4]. Some of the limitations of FUFM are that, the time factor is not considered in FUFM. Also it generates candidate itemsets to find FIS, which is a time consuming process. Further FUFM deals with all transactions of the database regardless of their utility value. But transactions with low utility value may not give any fruitful results. These issues are addressed in this paper using experimental results.

3 Proposed System Design and Methodology

The methodology for the proposed query driven system is shown in Fig. 4.

3.1 Construction of Itemset Tree with Utility

The algorithm CONSTRUCT_UTY shown in Table 3 is used to construct an itemset tree with utility.

The algorithm CONSTRUCT_UTY is applied on a sample database given in Table 1 and Table 2. The itemset tree is constructed as shown in Fig. 5. In Fig. 5, the itemsets are shown in solid boxes and the corresponding utility values are shown in dashed boxes. The following steps are carried out:

1. The first transaction is turned into the root, i.e., $T_R = \{1,2,5\}$, $f(T_R) = 1$ with utility {2,1,4} and NODE = 1
2. The second transaction, {1,2,4} is compared with the root $T_R = \{1,2,5\}$ and the largest common ancestor is {1,2} (case 3). Therefore create a new root L, where L = {1,2}and $f(L) = f(T_R) + 1 = 2$, NODE = 2. The current root R, and a new node {1,2,4} with f{1,2,4} = 1 and NODE = 3 are made the children of L. The utility values of L and the new node are updated as shown in the Fig. 5.
3. The third transaction, {1,3} is compared with the root $T_R = \{1,2\}$, and the largest common ancestor is {1} (case 3). Therefore create a new root L, where L = {1}and $f(L) = f(T_R) + 1 = 3$ with NODE = 4. The current root R, and a new node {1,3} with f{1,3} = 1 and NODE = 5 are made the children of L. The utility values of L and the new node are updated as shown in the Fig. 5.

In a similar way, the fourth basket {1,2,3,5} and the fifth basket {1,2,3} can be inserted and the resultant tree is as shown in Fig. 5.

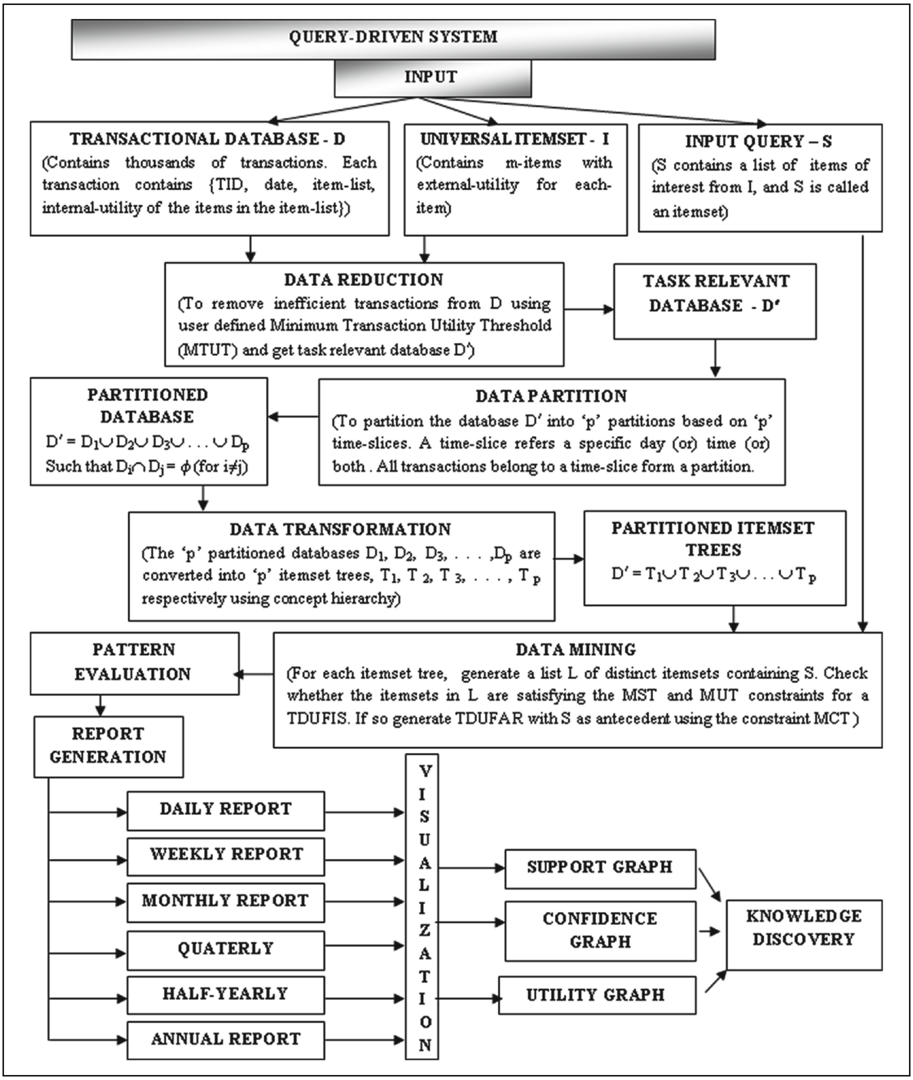


Fig. 4. Query driven system for finding Time-Dependent Utility Frequent Itemsets and Association Rules

Table 3. CONSTRUCT_UTY algorithm

```
Algorithm 1: CONSTRUCT_UTY(S, T, INT_UTY, UTY_SET) : Incorporating a Market
Basket transaction S, along with Internal utility INT_UTY , in an itemset
tree, T
Input:    Market basket Transaction S, Internal utility INT_UTY
Output: Itemset Tree 'T' and Utility set 'UTY_SET'
Let R= [S_R, f(S_R)] denote the root of  T and let [C_i ,f(C_i)], be R's children.
Set the first market basket transaction as the root node with node address
NODE=1 and f(S_R) = 1. Start the algorithm by inserting the second market
basket transaction onwards.
1.  If [S = S_R ] then set f(S_R) = f(S_R ) + 1 , and Update  utility set
    UTY_SET, with Internal utility INT_UTY
2.  Else if (S is an ancestor of S_R) then
    a) Create a new root [S, f(S)= f(S_R ) + 1], NODE=NODE+1
    b) Make [S_R ,f(S_R )] the child of the new root
    c) Update utility set UTY_SET, with Internal utility INT_UTY
3.  Else if L is the common ancestor of S and T_R (such that  L ≠ S , L ≠ S_R )
    then
    a) Create a new root [L, f(L)= f(S_R ) + 1], NODE=NODE+1
    b) Append to L two children: [S, f(S)= 1] , NODE=NODE+1 and  [S_R ,f(S_R )]
    c) Update utility set UTY_SET, with Internal utility INT_UTY
 4. Else
       a)  Set f(S_R ) = f(S_R ) + 1
       b)  Select an action from the following list.
           ▪ If (S_R is the only common ancestor of  S with any child of R , or
             if the set of R's children is empty), then create a new leaf,
             [S, f(S)=1] with NODE=NODE+1 and make it a child of R.
           ▪ If (C_i = S) , then  f(C_i) = f(C_i ) + 1
           ▪ If (S is an ancestor of C_i ) then insert  [S, f(S)=f(C_i )+1] with
             NODE=NODE+1 between R and C_i
           ▪ If  (C_i  is  an  ancestor  of  S)  then  let  T(C_i  )  be   the  subtree
             rooted    at    [C_i    ,f(C_i)];    Call   CONSTRUCT-UTY(S,T(C_i),
               INT_UTY,UTY_SET)
           ▪  If (L is the largest common ancestor of C_i and(such that L ≠ S,
              L ≠ C_i ) and the itemset associated with R is an ancestor
              of L), then:
                • Disconnect [C_i ,f(C_i)] from R;
                • Create  a  new  node,  [L,f(L)=f(C_i )+1]  as  a  child  of  R  and
                  NODE=NODE+1;
                • Append  to  L  two  children  :  [S,  f(S)=1]  with  NODE=NODE+1
                  and    [C_i ,f(C_i)];

       c) Update utility set UTY_SET, with Internal utility INT_UTY
```

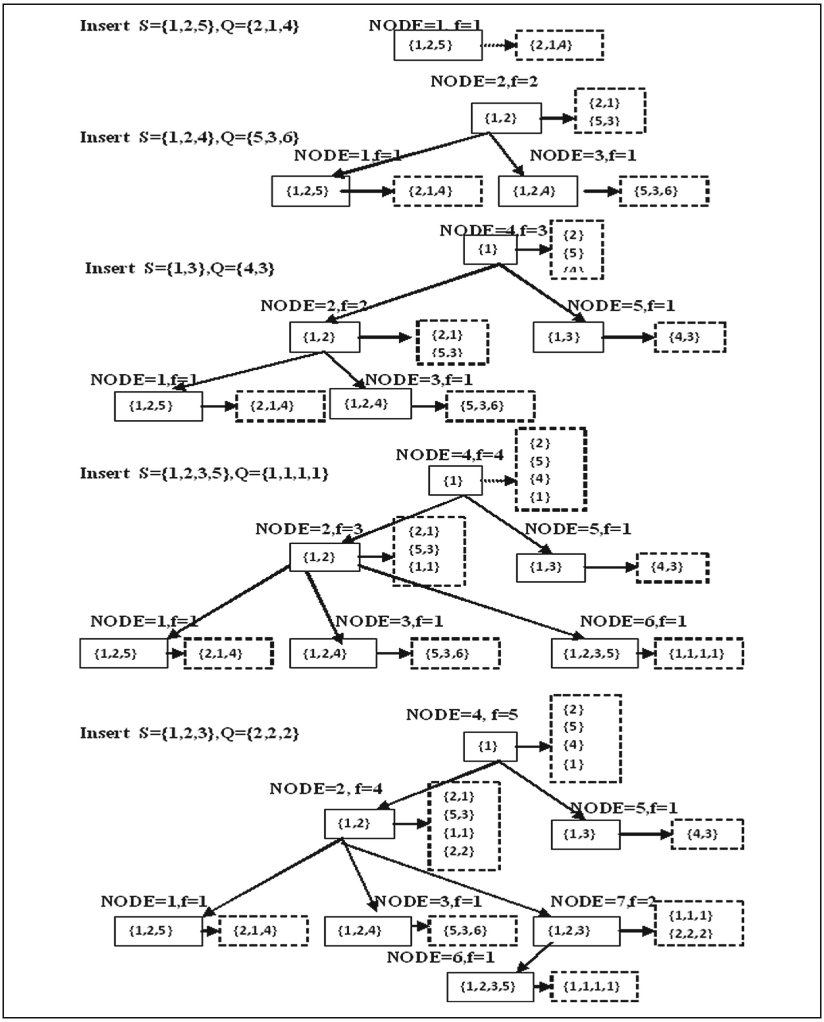


Fig. 5. Itemset tree construction from the set of market baskets in Table 1 and Table 2.

3.2 Support Count and Utility of an Itemset

The algorithm 2 in Table 4 is used to find the support count of an itemset S by traversing through the itemset tree constructed in Fig. 5.

Table 4. SUPPORTCOUNT_UTY algorithm

```
Algorithm 2: SUPPORTCOUNT_UTY (T,UTY_SET,S) - To find the Support Count and
Utility of an itemset S
Input   : Itemset tree T, Utility-Set UTY_SET, Itemset  S
Output : Support Count of S, SC(S) and Utility of S, u(S)
Notations:
R          – Root of Itemset Tree T
LCH (NODE) - Left Child of NODE
SIB (NODE) - Sibling of NODE
PAR (NODE) - Parent of NODE
DATA (NODE)- Data item of NODE
FC (NODE)  -  Frequent Count of NODE
SUPPORTCOUNT_UTY (T,UTY_SET,S)
1. NODE = R;
2. if (LCH(R) = 0) then return;
     else  NODE = LCH(R);
LOOP:
3. if (First-item(DATA(NODE)) > First-item (S)) then  goto CHECK-SIBLING;
4. if (S ⊆ DATA (NODE)) then
      {
          sc=sc+FC(NODE);
          FIND-UTILITY;
          goto CHECK-SIBLING;
      }
5. if (Last-item (DATA(NODE) > Last-item (S)) then goto CHECK-SIBLING;
6. if (LCH(NODE) = 0) then goto CHECK-SIBLING;
     else
           NODE= LCH(NODE);
                        goto LOOP;
7. CHECK-SIBLING:
     if (SIB(NODE)=0) then
     {
           if (PAR(NODE)=R) then goto DONE;
           else
  {
                  NODE = PAR(NODE);
                  goto CHECK-SIBLING;
           }
      }
      else
      {
  NODE=SIB(NODE);
             goto LOOP;
       }
8. Procedure FIND-UTILITY;
      freq = FC(NODE); ct=1; uty=0;
     Extract the Internal-Utility list corresponding to the NODE from the
     Utility set UTY_SET.
     while (ct ≤ freq)
     {   i=nitem
         uty=uty+ ∑(Ext-Uty of items in S) * (Int-Uty of items in S);
  i=1
  ct=ct+1;
     }
     return;
9. Procedure DONE;
  end;
```

Example. The details of DATA, FC, LCH, SIB and PAR for each NODE in the itemset tree in Fig. 5 are given in Table 5.

Table 5. DATA,FC,LCH,SIB and PAR of NODE

NODE	DATA(NODE)	FC(NODE)	LCH(NODE)	SIB(NODE)	PAR (NODE)
1	{1,2,5}	1	0	3	2
2	{1,2}	4	1	5	4
3	{1,2,4}	1	0	7	2
4	{1}	5	2	0	0
5	{1,3}	1	0	0	4
6	{1,2,3,5}	1	0	0	7
7	{1,2,3}	2	6	0	2

Let us calculate the support count and utility of itemset S = {1,3} using the itemset tree in Fig. 5 as detailed below.

The process starts with the root node (i.e.) NODE = R=4 and LCH(R) $\neq$ 0. The left most branch of the root node is examined. First it visits {1,2} and then {1,2,5}. Since Last-item({1,2,5}) > Last-item({1,3}, the control goes to check the sibling of {1,2,5}, which is {1,2,4}. Again, Last-item({1,2,4}) > Last-item({1,3}. Once again it checks the sibling of {1,2,4} which is {1,2,3}. Now S = {1,3} $\subseteq$ {1,2,3}. Hence sc = sc + FC (NODE). i.e., sc = 0+2 = 2. Also the utility value of {1,3} is calculated by extracting the respective internal utility values of the node {1,2,3}, which are ({1,1,1},{2,2,2}). The external utility values of {1,3} are {5,1} from Table 1. Therefore uty = uty + (1 + 2)*5 + (1 + 2)*1 = 15 + 3=18.

By continuing this process we can find a node {1,3} such that S = {1,3} $\subseteq$ {1,3}. The support count and utility are updated as sc = sc + FC(NODE). i.e., sc = 2+1 = 3 and uty = 18 + (4*5 + 3*1) = 41. Now there is no sibling for the node {1,3} and hence the control goes to its parent which is the root node and the process ends with SC {1,3} = 3 and u{1,3} = 41. The support count and utility share of the itemset S are calculated in terms of percentage as: SC(S) = SC(S)/n(D)*100 = 3/5*100 = 60 % and U (S) = u{S}/u(D)*100 = 129*100 = 32 %

4 Experiments

4.1 Description of the Medical Database

The technical terms related to the medical database are described below:

1. Itemset, I = $\{s_1, s_2, \ldots s_q, d_1, d_2, \ldots d_r\}$ refers ‘q’ symptoms $\{s_1, s_2, \ldots s_q\}$ and ‘r’ diseases $\{d_1, d_2, \ldots d_r\}$. In the case study, a sample of 20 symptoms and 15

diseases are taken as shown in Tables 6 and 7. The external utility EXT-UTY of the disease and symptoms refer to their level of seriousness. High EXT-UTY refers to high risk and low EXT-UTY refers to low risk.

Table 6. Symptoms codes and their intensities (external utilities)

S.No.	SYMPTOMS	SYM-CODE EXT-UTY(SYM) = (SYM-CODE % 3) + 1		
		LOW(01)	MEDIUM(02)	HIGH(03)
1	FEVER	21	22	23
2	BODY-PAIN	24	25	26
3	SUGAR	27	28	29
4	B.P	30	31	32
5	COLD	33	34	35
6	HEADACHE	36	37	38
7	COUGH	39	40	41
8	SHIVERING	42	43	44
9	VOMITING	45	46	47
10	LOOSE MOTION	48	49	50
11	BACK-PAIN	51	52	53
12	EYE-INFECTION	54	55	56
13	THROAT PAIN	57	58	59
14	STOMACH-PAIN	60	61	62
15	CHEST PAIN	63	64	65
16	URINARY INFECTION	66	67	68
17	INDIGESTION	69	70	71
18	KNEE PAIN	72	73	74
19	NERVOUSNESS	75	76	77
20	MOTION PROBLEM	78	79	80

2. It is assumed that each record in the medical database D contains the following fields:

PID(Patient-ID): For ‘n’ patients, the PID’s are PID_1, PID_2, PID_3,. . ., PID_n

TIME: It may refer to the patient infection date (or) time (or) both. Here it is assumed to be the date in the range [1. .365], corresponds to 365 days in a year. (i.e.,) 1 – JAN’1, 2 - JAN’2, . . . 365 - DEC’31.

Table 7. Disease-type with external utility and possible symptoms

DIS CODE	Name of the disease	EXTUTY	Three Important Symptoms (Note: The actual symptoms may vary. The number and order of the symptoms may also vary)		
			SYM-1	SYM-2	SYM-3
01	VIRAL-FEVER	1	FEVER	BODY-PAIN	SHIVERING
02	DIARRHOEA	2	LOOSE-MOTION	VOMITING	STOMACH PAIN
03	SINUS	3	HEADACHE	COLD	COUGH
04	APPENDICITIS	4	STOMACH PAIN	INDIGESTION	MOTION PROBLEM
05	STONE IN KIDNEY	5	URIN-INFECTION	STOMACH PAIN	INDIGESTION
06	EYE-SIGHT DEFICIENCY	6	SUGAR	HEADACHE	NERVOUSNESS
07	PILES	7	MOTION PROBLEM	STOMACH PAIN	INDIGESTION
08	ASTHMA	8	COUGH	COLD	CHEST-PAIN
09	PNEUMONIA	9	FEVER	COUGH	COLD
10	MALARIA	10	FEVER	SHIVERING	BODY-PAIN
11	TYPHOID	11	FEVER	BODY-PAIN	HEADACHE
12	FITS	12	FEVER	NERVOUSNESS	B.P
13	JAUNDICE	13	URINARY-INFECTION	NERVOUSNESS	SUGAR
14	DENGUE FEVER	14	FEVER	BODY-PAIN	SHIVERING
15	A1H1 - SWINE FLU	15	COUGH	FEVER	THROAT PAIN

DIS-SYM1-SYM2-SYM3: It refers to the disease-symptom code of the patient. DIS denotes the disease code and SYM1, SYM2 and SYM3 denote three important symptom codes for the disease. It is assumed that the EXT-UTY(DIS) = DIS, and EXT-UTY(SYM) = (SYM % 3) + 1, SYM may refer SYM1 or SYM2 or SYM3. For example, if DIS-SYM = 01-212434. Then the EXT-UTY will be 01-010102, which is calculated as shown in Table 8.

Table 8. Calculation of EXT-UTY from DIS-SYM

	CODE	DESCRIPTION	EXT-UTY
DIS	01	VIRAL-FEVER	01
SYM1	21	LOW FEVER	(21 %3) + 1=01
SYM2	24	LOW BODY-PAIN	(24 %3) + 1=01
SYM3	34	MEDIUM COLD	(34 %3) + 1=02

PRE-HIS: The previous case history of a patient serves as the internal utility (INT-UTY) of the medical transactions. INT-UTY gives the frequency of the symptoms and diseases occurred earlier. For example, if for a patient, PRE-HIS="02-010201" and DIS-SYM="01-212434". It means that the patient got affected earlier with the disease VIRAL-FEVER two times, the symptoms LOW-FEVER one time, LOW BODY-PAIN two times and MEDIUM-COLD one time. The utility value of this transaction is got by the sum of the products of INT-UTY and EXT-UTY (i.e.,) u(T) = 1*2 + 1*1 + 1*2 + 2*1 = 7.

3. In generating association rules of the form A ⇒ B, It is assumed that the antecedent A = { s_1, s_2, . . . s_i} is a set of 'i' symptoms among 'q' symptoms and B is a set of only one disease among 'r' diseases. The rules are validated using three measures viz., confidence, support count and utility with respect to the time.

4.2 Application of TD-FVAUFM on the Medical Database

The following steps are carried out on implementing TD-FVAUFM on the medical database D.

1. The medical database D is generated using a sample data collected from a popular hospital in the city. The database D contains 26258 transactions for one year.
2. The transaction utility u(T), ∀ T ∈ D and the maximum utility of the transaction is calculated. Here, M = MAX { u(T), T∈D } = 55 and u(D) = 384069.
3. Let the Minimum Transaction Utility Threshold (MTUT) = *min_trn_uty* = λ = 15 % A transaction is said to be an efficient transaction if it satisfies the constraint u(T) / M ≥ λ = 15 % (or) u(T) ≥ λ * M = 15 % of 55 ≅ 8.
4. Perform **data reduction** on the database D, by removing the inefficient transactions T from D, for which u(T) < 8 and construct the new task-relevant database say D′. It is found that D′ contains 21604 transactions and u(D′) = 359687.

5. Apply **data partition** to divide D′ into 365 partitions corresponding to 365 days.
6. Apply **data transformation** technique to transform the database partitions in the form of an itemset tree using the CONSTRUCT_UTY algorithm.
7. The original database D is now in the form of 365 itemset trees along with the corresponding utility set. Given a query S, TD-FVAUFM generates association rules with S, as antecedent. The graph structure is used to visualize the rules.

4.3 Result Analysis of TD-FVAUFM

Let us analyze the response for the query S = {22,24}, the case where there is medium fever and low body-pain. The association rules for the query S are generated as shown in Table 9 and multi-dimensional reports on association rules are shown in Table 10.

Table 9. List of Association Rules

ASSN.RULE	MEANING
{22,24} ⇒ {01}	S ⇒ {Viral-Fever}
{22,24} ⇒ {10}	S ⇒ {Malaria}
{22,24} ⇒ {11}	S ⇒ {Typhoid}
{22,24} ⇒ {14}	S ⇒ {Dengue-Fever}
{22,24} ⇒ {15}	S ⇒ {Swine-Flu}

The following observations are made from Table 10

1. In Table 10, the various notations refer to: c - *min_conf,* α - *min_sup*, μ - *min_uty*, * - *low frequency*, # - *low utility*,@ - *low confidence*.
2. Day-wise analysis lists association rules with 100 % confidence. Most of the Patients are affected with Viral-Fever and Typhoid than other diseases. More over Typhoid and Malaria has large utility value that shows the seriousness of those diseases. Dengue Fever and Swine Flu occurs less frequently but with large utility (i.e.,) they are dangerous diseases but occur less frequently.
3. In week-wise analysis the confidence level reduces to 50 %. Typhoid affects in 11 weeks with high support counts and utility. Viral-Fever affects in 7 weeks with normal support count and low utility. Malaria gets affected in 4 weeks with large support counts and utility.
4. In Monthly analysis, Typhoid is the only disease that has significant impact in the months of Jun and Dec with less confidence in the months of Jan, Feb and Aug.
5. In Qtr-Year, Half-Year and Yearly analysis the confidence level for Typhoid is around 35 to 40 %.

Table 10. TDUFAR for the query S = {22,24}

DAY-WISE ANALYSIS (c=100, s=3, μ=1)				WEEK-WISE ANALYSIS (c=50 , s=1.0 , μ=0.4)			
S ⇒ {01-VIRAL -FEVER}				S ⇒ {01-VIRAL - FEVER}			
Day	**Con**	**Sup**	**Uty**	**Week**	**Con**	**Sup**	**Uty**
12	100	3.2	1.1	2	50	1.2	0.4
95	100	3.0	1.0	13	50	1.2	0.4
114	100	3.3	1.2	17	70	1.7	0.6
149	100	3.1	1.0	41	50	1.2	0.4
162	100	3.2	1.2	42	60	1.4	0.4
247	100	3.2	1.0	47	63	1.2	0.4
S ⇒ {10 – MALARIA}				48	50	1.4	0.5
14	100	3.4	4.5	S ⇒ {10-MALARIA}			
255	100	3.2	3.3	8	50	1.4	1.4
280	100	3.1	3.2	10	64	1.6	1.8
329	100	3.3	3.7	20	67	1.4	1.3
S ⇒ {11 – TYPHOID}				40	60	1.4	1.6
2	100	3.3	3.0	S ⇒ {11-TYPHOID}			
10	100	3.2	3.7	1	50	1.4	1.2
180	100	2.9	2.4	3	50	1.7	1.9
229	100	3.1	3.3	6	71	1.2	1.3
232	100	3.0	2.6	16	63	1.6	1.6
307	100	3.2	3.8	25	50	1.2	1.0
355	100	3.0	2.6	26	69	2.1	2.0
S⇒{14 –DENGUE-FEVER}				34	56	1.2	1.3
148	100	1.7*	3.1	44	57	1.0	1.0
172	100	1.6*	2.9	49	50	1.7	1.7
293	100	1.6*	1.5	51	73	1.9	1.7
344	100	1.9*	1.8	52	67	1.4	1.3
S ⇒ {15 – SWINE FLU}				MONTHLY ANALYSIS (c=50, s=1, μ=0.5)			
242	100	1.6*	3.2	S ⇒ {01-VIRAL -FEVER}			
317	100	1.7*	1.6	**Mon**	**Con**	**Sup**	**Uty**
QTR-YR ANALYSIS (c=35, s=0.5, μ=0.5)				9	49@	1.0	0.3#
S ⇒ {11-TYPHOID}				S ⇒ {11-TYPHOID}			
Qtr	**Con**	**Sup**	**Uty**	1	42@	1.0	1.0
1	36	0.8	0.9	2	40@	0.8*	0.9
2	37	0.8	0.8	6	56	1.2	1.1
3	35	0.7	0.8	8	43@	0.8*	0.9
4	41	1.0	1.0	12	50	1.4	1.4
HALF-YR ANALYSIS (c=35, s=0.5, μ=0.5)				YEARLY ANALYSIS (c=35, s=0.5, μ=0.5)			
S ⇒ {11-TYPHOID}				S ⇒ {11-TYPHOID}			
Half	**Con**	**Sup**	**Uty**	**Year**	**Con**	**Sup**	**Uty**
1	36	0.8	0.8	1	37	0.8	0.9
2	38	0.9	0.9				

4.4 Visualization of Association Rules with Confidence, Support and Utility

From the confidence graph in Fig. 6, it is observed that there are more chances of getting affected with the diseases Typhoid, Malaria & Viral Fever than the diseases Dengue fever and A1H1 Swine Flu. Also Typhoid occurs in the month of June (6th month) with 56 % confidence.

The support count graph in Fig. 7 shows that the diseases Typhoid and Viral-Fever have great impact over other diseases. The maximum support count for the year lies in the month of December with 1.44 % for the disease Typhoid.

The utility graph in Fig. 8 shows that the disease Typhoid has a great impact over other diseases and the maximum utility lies in December with 1.39 %.

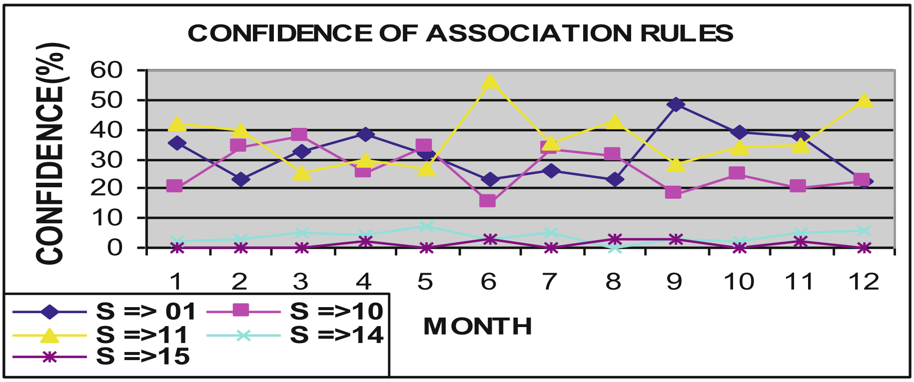


Fig. 6. Month-wise Confidence graph of association rules for the query S = {22,24}

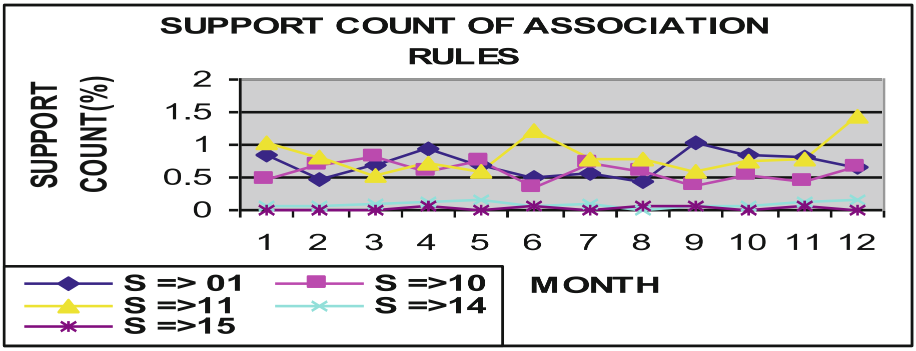


Fig. 7. Month-wise Support Count graph of assn. Rules for the query S = {22,24}

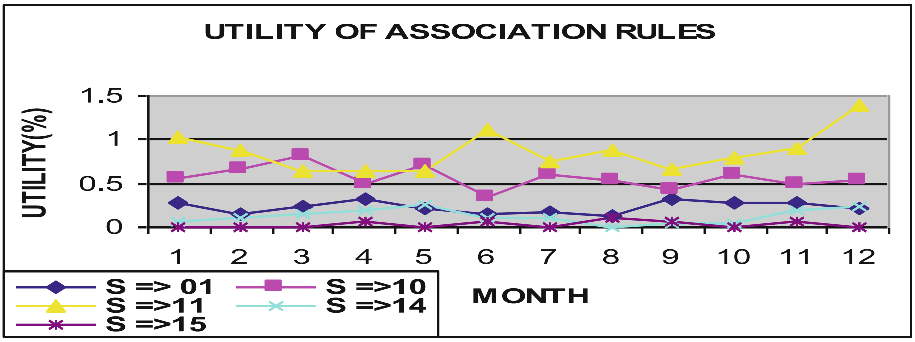


Fig. 8. Month-wise Utility graph of association rules for the query S = {22,24}

5 Conclusion and Suggestions for Future Research

In this paper, a query driven system is proposed for generating association rules based on the time and utility factors. The proposed system is applied on a medical database. Given a query in the form of symptoms of a patient, the system is able to predict the disease type by generating time-dependent association rules. The rules are validated with three measures viz., confidence, support count and utility.

The proposed system can be viewed as a mathematical model and can be applied on any type of databases. The application of the system depends on the creativity of an individual. The key aspect in applying the system on a database is to define:

The items used in the database, the internal utility of the items in each transaction of the database and the external utility of the items of the database.

The system proposed in this paper can be further extended to generate multi-level association rules. Also in the medical database the clinical results of a patient may be included for making better predictions.

Acknowledgments. The encouragement and financial support provided by the Sourashtra College authorities during the preparation of this research work is highly acknowledged.

References

1. Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large databases. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 207–216. Washington, May 1993
2. Arumugam, G., Vijayakumar, V.K.: Discovery of time-dependent association rules for targeted queries on medical data. In: Proceedings of the International Conference on Recent trends in Information Systems (IRIS 2006) National Engineering College, Kovilpatti, Tamil Nadu, India, pp. 443–453, 6–8 January 2006
3. Borgelt, C.: Keeping things simple: finding frequent itemsets by recursive elimination. In: Workshop Open Source Data Mining Software, pp. 66–70. ACM Press, NewYork (2005)
4. Han, J., Pei, J., Yin, Y.: Mining frequent patterns without candidate generation. In: Proceedings of the International Conference on Management of Data, pp. 1–12, 2000
5. Han, J., Kamber, M.: Data Mining – Concepts and Techniques. Morgan Kaufmann Publishers, San Francisco (2001)
6. Kubat, M., Hafez, A., Raghavan, V.V., Lekkala, J.R., Kian Chen, W.: Itemset Trees for Targeted Association Querying. IEEE Trans. Knowl. Data Eng. **15**(6), 1522–1534 (2003)
7. Adnan, M.H.M., Husain, W., Rashid, N.A.: Data mining for medical systems: a review. In: Proceeding of the International Conference on Advances in Computer and Information Technology – ACIT, pp. 17–22 (2012)
8. Podpe.can V.: Utility-based Data Mining. BSc Paper (in Slovene), University of Ljubljana (2007)
9. Podpecan, V., Lavrac, N., Kononenko, I.: A Fast Algorithm for Mining Utility-Frequent Itemsets (2007)
10. Ruijuan, Hu: Medical data mining based on association rules. J. Comput. Inform. Sci. **3**(4), 104–108 (2010)
11. Doddi, S., Marathe, A., Ravi, S.S.: Discovery of association rules in medical data. Med Inform Internet Med. **26**(1), 25–33 (2001)
12. ACM SIGKDD Workshop on utility-based data mining, (2005). http://storm.cis.fordham.edu/˜gweiss/ubdm-kdd05.html
13. ACM SIGKDD Workshop on utility-based data mining (2006). http://www.ic.uff.br/bianca/ubdm-kdd06.html

14. Weiss, G., Zadrozny, B., Saar-Tsechansky, M.: Utility-baseddatamining 2006 workshop report. SIGKDD Explorations **8**(2), 98–101 (2006)
15. Yao, H., Hamilton, H.J., Butz C.J.: A foundational approach to mining itemset utilities from databases. In: The Fourth SIAM International Conference od Data Mining SDM, pp. 428–486 (2004)
16. Yao, H., Hamilton, H.J., Geng, L.: A unified framework for utility based measures for mining itemsets. In: Second International Workshop on Utility-Based Data Mining, Philadelphia (2006)

Cord Blood Collection Promotion Through Knowledge Sharing Using Change Management HSE Model

Fatma Alhashimi(✉)

The British University in Dubai BUID, Block 11, 1st and 2nd Floor, Dubai International Academic City, PO Box 345015, Dubai, United Arab Emirates
alhashimi82@hotmail.com

Abstract. Despite the high level of births in Dubai, there are a low number of cord blood donations at ABC- Cord Blood center. A survey was conducted to identify the problem. The survey shows that a lack of knowledge is one of the major barriers that stopped expectant mothers from donating their baby's umbilical cord blood. Knowledge sharing is vital to promote awareness among expectant mothers about the importance of cord blood stem cells which in return will increase number of cord blood donations in the center. This paper describes new approach using a change management model HSE (Health Services Executive) to promote knowledge sharing and awareness among expectant mothers. The number of collected private and public cord blood units has increased significantly by 57 % and 300 % respectively which demonstrated the effectiveness of change management in promoting knowledge awareness and knowledge sharing.

Keywords: Change management · Knowledge awareness · Knowledge sharing

1 Introduction

Many people in the UAE are not aware about the importance of cord blood donation and its impact on the community. A survey was distributed to 65 expectant mothers in two largest government hospitals in the UAE. Results showed that most of the expectant mothers do not donate because they do not have knowledge about cord blood stem cells to help them make an informed decision. It is therefore important to increase knowledge sharing and awareness among relevant stakeholders. Stakeholders in this project include the author, healthcare educators, donors, marketing department and the management in the maternity hospitals.

Healthcare delivery is defined as a knowledge driven-process leading to knowledge management (KM). Knowledge management provides a great chance for performance development [1]. Reference [2] Showed that finding, sharing and increasing knowledge are considered to be the main critical factors of the KM process, thus leading to increased organizational performance and productivity. During the diagnostic-therapeutic cycle, healthcare knowledge plays a vital role in clinical decision making. It is important that knowledge is derived and applied within correct diagnostic decisions so that it leads to

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 683–694, 2015.
DOI: 10.1007/978-3-319-21009-4_50

effective therapeutic management [2]. Reference [3] Defined Healthcare Knowledge Management (HKM) as organized modeling and sharing of healthcare knowledge to develop high quality patient care. These authors argue that the main objective of HKM is to deliver healthcare knowledge in timely and optimal manner to healthcare professionals, individuals and patients, which in turn will lead to the achievement of high quality service and cost effective patients' care decisions. We agree with [3] that the significant aim of HKM is promoting healthcare delivery system where healthcare knowledge is considered an essential resource in improving health outcomes.

It is our belief that in order to overcome the cord blood donation problem in UAE, we need to take a HKM approach. However, the HKM portfolio still presents many challenges, specifically in the development of knowledge-centric services which seamlessly incorporate within the population [3]. There are multiple levels of challenges such as the design of health services that can provide value and meet the expectant mothers' needs as well as the utilization of service by stakeholders. It is our belief that knowledge sharing plays a vital role in solving the problem. This paper describes how the use of knowledge management through change management can be used to increase the awareness of the expectant mothers in the cord blood donation problem in UAE. The paper begins with a brief review of knowledge sharing, followed by change management. This is then followed by discussions of the HSE model that we used to improve the experience of donors and the healthcare users [4]. Evaluation of the results is briefly reviewed in subsequent sections. The paper concludes with suggestions for further research.

2 Knowledge Sharing

The objective of all health care is to enhance the quality of care to all patients. Consequently, it is important to find, to share, to collaborate, and develop clinicians' knowledge and quality of care. Researchers argued that adoption of knowledge management techniques provides the mean to improve the quality of care [2].

Knowledge is an important intangible asset. According to [5], Knowledge management is concerned with the creation, use, reuse and dissemination of knowledge. It is also the process of creating value from an organization's intangible assets [5]. Reference [6] Points out that Knowledge management not only involves the production of information, but it also allows us to capture the data at the source, to transmit and analyze it, as well as to communicate information [6]. Successful implementation of KM can lead to improved knowledge-sharing within the organization [7].

Knowledge sharing (KS) has become a crucial component of knowledge management (KM). Researchers such as [8] described knowledge sharing as the delivery of task information that plays important role in knowing how to solve problems and collaborate with each other which will result in developing new ideas and creating new policies and procedures [8]. Another concept of Knowledge sharing involves the processes through which knowledge is channeled between a source and a recipient. The aim of any knowledge-sharing process is to transfer source knowledge to a recipient successfully. Another way of defining knowledge sharing is how knowledge is re-created in the recipient.

Reference [9] Argued that Knowledge sharing is about the willingness of individuals within a team to share with others the knowledge they have acquired [9]. Choi and his colleagues suggested that knowledge sharing is a multidimensional activity. It involves several contextual, cognitive, and communicative skills. [10].

Knowledge sharing can happen through written communication such as emails and post or face-to-face communications via networking with other experts. It also occurs via recording, organizing and capturing knowledge for others. There are several ways to share knowledge. The most popular way is through meetings. In meetings people come together formally to discuss about a problem, or a projects. It is also where people meet to express opinions and make decisions. Video Screening Sessions can be used as an effective means in KS. During the video screening sessions, participants share a film (movie) or television program with others. The program normally contains past experiences, case studies, best practices and methods used by the participants.

Another method for knowledge sharing is the use of Training sessions /training programs. In these sessions, participants learn the skills, procedures, activities for a particular job or task. During the Training programs the trainees cooperate with each other and interact with other trainers/professionals [9]. In the session, the trainer shares his knowledge with the trainees and trains them to do a particular job. Besides these, other techniques include: Peer assist; Retrospective Review; Story–telling/scenarios; Narrative case study; after action review; Mentoring; Knowledge Café; Web 2.0 – Social networking etc.

Although each of these has its benefits for sharing knowledge, we decided to adopt a change agent approach using change management. It is our belief that knowledge sharing among the different stakeholders requires the change of daily routine, behavior, and often the processes and organizational structures in the cord blood project. Therefore, it is important to take the theories of change management into consideration. In order to promote a change management approach for the project. We have decided to use the HSE for Change Management.

3 Change Management

Change management is the approach that plans, maps, applies, controls, determines, and maintains changes in work and business processes [11]. By reviewing change management literature, there is significant disagreement surrounding the most appropriate and best approach to changing organizations [12]. There are two main approaches to change, which are 'planned' and 'emergent' [13]. There are different change models to design, plan and implement successful change.

The Planned Approach. The planned change approach has dominated for the past 50 years and it is based mainly on the work of Lewin. His theory focuses on moving from one fixed state to another through different pre-planned steps, such as Lewin's "action research" model and Lewin's "three step model" which illustrates the main three learning stages of freezing [13]. Lewin's theory is a planned (prescriptive) approach that mainly focuses on finding an effective approach to resolving social conflict by understanding and changing group behaviour and tries to map out the difficulty of the field in which the behaviour takes place [14]. The concept of planned change illustrates that an organization is present in diverse states at different times and there is a planned movement which occurs from one state to another. Although the Planned change approach is highly effective, there are many criticisms against it. Firstly, it is based on the assumption that an organization operates under stable conditions, and by a pre-planned manner it can move from one stable state to another. However, organizational change is defined as being a continuous and open-ended process. Secondly, such change is not suitable for situations that require rapid transformational change [15]. Thirdly, the need of learning new major methods in a new situation as it is not clear when the new refrozen state will be reached. Moreover, this type of change approach model mainly depends on employee concerns and behaviour which is unsuitable in some other circumstance and in organizational conflicts and politics [14].

Emergent Approach. However, a new concept of change called "emergent approach" which is used to achieve better understanding of the problems to be able to manage change within the developed complex environment. The concept of emergent approach of change is different from planned change. It consists of on-going accommodations, adaptation and alteration to achieve fundamental change [14]. Kotter's change model is an example of an emergent approach. This approach is able to achieve better understanding of the problems of managing change within a complex environment [15]. There are no specific general rules for emergent change. It mainly focuses on five main features of organizations which are: structures, cultures, organizational learning, managerial behaviour and politics. The emergent model views change as a driven process from bottom up instead of top down and also views change as an open-ended and continuous process of learning. Kotter's model of change prescribes a linear approach to change by following a number of steps in an ordered and logical sequence [16]. The key eight steps of Kotter's model provide a path to success. Changing the behaviour of people is considered the most fundamental problem in all of the stages [18]. It is very important to help people see and feel the change in order to achieve successful change [19]. Kotter's eight steps can be divided into three main categories which are preparation, action and grounding [20].The limitation of the emergent approach is that it is a less comprehensive approach and it is limited in different types of organizational change.

Organizational Development Approach. The third approach is the Organizational development approach. An example of this approach is the HSE (Health Services Executive) change model. HSE change model is new organizational development approach, which been developed to improve the experience of patients and services, encourage team working and promote constant approach to change across the system [21]. HSE focuses on people and cultural aspects of change, as managing organizational change is all about managing people. Thus, organizations should not ignore or discount the employee perspective [22]. The main important function of leaders is to manage their own culture by understanding individuals' beliefs, thoughts and feelings [23]. HSE reflects and builds on specific core principles that can be applied by leaders and managers. It is based mainly based on four main stages of project management lifecycle, which are: initiation, planning, implementation and mainstreaming. All four stages of the HSE model are interrelated and influence each other [4].

Critical Review About Change Management Models. A number of models were reviewed. Lewin's model is considered one of the core models of change. His theory mainly tries to understand how social groupings are formed, motivated and maintained. Although Lewin's concept is straightforward and simple, it was not selected by the author due to the fact it is too restrictive and prescriptive because it does not pay enough attention to the need to analyse and conceptualize organizational change [24]. Although planned change has proved to be useful under stable conditions, it does not make sense to implement planned processes for freezing change behaviors which are complex and dynamic processes that cannot be solidified and treated as a series of linear sequence [13]. Furthermore, the three stages of change by Lewin's model provide no option for the author to discuss the importance of knowledge and awareness during the change. Lewin's model was mainly designed for stable organizations which do not accommodate the multiple changing environments found within the ABC-cord blood center.

The author also viewed the 8 steps of Kotter's change model where all the steps are structured and organized. Despite the author favouring this model, it was not selected due the fact that in order to use this model successfully, steps should be implemented selectively and adaptively in order to best match the culture and environment of the organization. Kotter's change model lacks flexibility that allows the author to go back to the previous steps. It was not suitable to be applied to the ABC-cord blood center where the environment and conditions require flexibility. For ABC-cord blood center to be successful, it is very important to choose the right model so that the chosen model has the potential to contribute to achieving successful change outcome [16].

Therefore, after studying the different change models and the conditions and environment of the ABC-cord blood center, the author found that the HSE change model was the most appropriate as it met the project requirements and thus allows the flexibility to move back to the previous steps. Another reason for choosing the HSE change model was because of the continual cyclical nature of the model [17]. This model also covers most of the issues that the author wants to address. It concentrates on the importance of engagement, communication and decision making to continue organizational improvement. As our research was to promote knowledge sharing and awareness about the importance of cord blood stem cells, it is important for us to be able to communicate and increase knowledge and learning between expectant mothers by choosing the right communication style, which is discussed in HSE model.

4 HSE Change Model for ABC-Cord Blood Center

This section describes the proposed change management approach HSE (Health Services Executive) which was used in ABC-Cord Blood Center.

ABC Case Study. ABC – is a cord blood center – that provides private and public banking of cord blood stem cells for donors. It consists of two main units which are administration and laboratory. The administration unit provides registration services, information and instructions for expectant mothers on cord blood and stem cells technology. The laboratory unit processes, tests, and stores the baby's umbilical cord blood for thirty years in liquid nitrogen tanks for transplantation purposes. Banking cord blood stem cells is considered as one of the latest new technologies that should be given attention and consideration in the UAE. The survey was distributed to expectant mothers in two government hospitals to find out what stops expectant mothers from donating their baby's umbilical cord blood stem cells and what are the factors that act as a barrier towards donation. Understanding blood donors' barriers is essential to improve the effectiveness of knowledge sharing and awareness.

The HSE change model consists of four main stages of change which are initiation, planning, mainstreaming and implementation. Each stage consists of important sub-stages that can help lead the change.

4.1 Initiation

The main purpose of the initiation stage is to create readiness to establish a sense of shared responsibility by building core leadership and management responsibilities [4]. This stage also covers scoping and planning to initiate the change and the selection of the stakeholders involved within the change. At this stage, it is very important to identify the main driving forces and the need to create the change by studying SWOT analysis which includes: Strengths, Weaknesses, Opportunity and Threats [25].

Preparing to Lead the Change. The purpose of this step is to build a foundation for effective change and develop a strong business case for successful change and that was done by conducting the following points:

- Identification of the driving forces and the need for change.
- Getting the authorization from higher management.
- Leadership roles and the key stakeholders.
- Identification of the driving forces and the need for change.
  SWOT analysis was conducted on the CBC (Cord Blood Center) to find out the main strengths, weaknesses, opportunity and threats. In the CBC, the main force that drives change is the number of cord blood units (private and public) received per year is low. Therefore, it was very important to identify the reason behind this low number of units and to increase the sense of urgency and clarify the purpose of change among relevant stakeholders to obtain their commitment and prepare them to lead the change.
- Getting the authorization from higher management.
  Despite the importance of cord blood stem cells, the author noticed that ABC-cord blood center was receiving on an average of 4–6 units per week. The issue was then discussed with higher management who suggested studying the problem and finding solutions to improve the number of units received. However, before the meeting date, the writer decided to carry out a survey to find out why the number of cord blood units received was low in the past years. The questions of the survey were prepared and approved by a quality department; the author then distributed the survey to 65 expectant mothers to test the level of knowledge about stem cells. Fifty percent (50 %) of them have very little knowledge and don't donate their cord blood because they do not have enough information to make a decision. This was one of

the most important findings to share with higher management to seek their approval to establish an educational unit that would carry out different functions such as informative cord blood presentations, awareness campaigns to promote knowledge sharing and awareness among expectant mothers about the importance of cord blood stem cells.

- Leadership roles and the key stakeholders.
  It is very important to identify the main stakeholders and establish an effective team to support the change [4]. The author conducted a short presentation illustrating figures that present low numbers of cord blood units received in the past years. To increase the sense of urgency, the author also shared with the staff and management the results obtained from the survey. The main purpose was to make them feel the need for change and thus provide help and support to develop as a successful team.

4.2 Planning

The main purpose of the second stage is to critically get the support and the engagement of staff and the key stakeholders. At this stage the author studied how to promote knowledge sharing and increase awareness among expectant mothers and make sure all the plans are decided and shared with the stakeholders. This was carried out by building commitment, determining the details for change and developing the implementation plans.

4.2.1 Building Commitment

The author was aware that with an increased number of UCB units, the workload will be increased and staff might resist the change. Therefore, building commitment and continued communication with the staff was very important to ensure that stakeholders will continue supporting the change until the end. This was obtained by assuring them of the beneficial impact of the change on the organization, building shared vision and communicating the vision to the business case of the change.

4.2.2 Determining the Details for Change

The purpose of this step is to assess the current situation so that the details for change can be determined. Assessing the current situation and analysing the information between the present situation and the shared vision of the future will help provide more information that supports the change [17]. In the CBC, the statistics from the previous years were obtained to see the number of units collected per year by the centre. It was identified that the number of cord blood units was low and change needed to occur to promote knowledge sharing and thus increase the number of collected units. Establishing an education unit will help spread awareness and knowledge sharing among expectant mothers. By comparing the statistical information from the past years the problem was highlighted and change was required to obtain the desired objective.

4.2.3 Developing the Implementation Plans

At this stage, the requirement is to allocate the details of implementation/project plans to bring together all the necessary details that will be conducted during implantation [4]. Different planned actions were conducted such as educational presentation in different hospitals, awareness campaigns, informative road shows about cord blood, designing brochures and posters, connecting with media for spreading awareness in news and TV channels, creating video and documentary for CBC.

4.3 Implementation

Implementing the Change. The main purpose of this stage is to implement the agreed changes and actions by supporting the staff to adopt new skills for changes. In September 2012, implementation took place at CBC; the education team were fully trained on how to conduct successful awareness to increase knowledge about the importance of cord blood. The tools were used as follows:

- Conducting presentations about the importance of cord blood banking in different maternity hospitals. The presentation was conducted among expectant mothers in ante-natal clinics.
- Designing brochures, roll up, posters and educational video documentary.
- Media coverage.
- Conducting road -shows in women's associations and ladies clubs.

4.4 Mainstreaming

The main purpose of this step is to help people practice new behaviours and skills and to learn from the way the change was processed to seek for continues development [4]. At this stage, the staff needed to feel and sense the achievement. Therefore, it is very important to acknowledge staff for their hard work and achievement [4].

4.4.1 Acknowledge Success and Achievement

ABC-cord blood center performed an award ceremony in January 2013 to reward nurses and gynaecologist doctors who collected best quantity and quality of cord blood stem cells. The staff working in the centre were also awarded for their hard work and support to increase awareness. Special awards and gifts were given by the director for each of the staff, doctors and nurses who contribute to the success of this project.

4.4.2 Evaluating and Learning

The main purpose of this step is to learn from the way the change was processed to seek for continuous development [4]. Many lessons were learned during the implementation stage of this project, one of the main points is to continue follow up at every stage of change to make sure tasks are done properly and on time. Another important point is to assign people to different tasks rather than depending on one person. During change implementation, the author was carrying out different activities at the same time.

It would have made the job easier to assign different staff to be responsible on each task so that the author followed up some important problems/tasks that appeared during implementation stage.

5 Evaluation of the Project

Evaluation of the changes was analysed using quantitative method which includes pre and post data collection of the number of collected UCB Umbilical cord Blood) units before and after carrying out educational awareness activities. The number of collected UCB units were gathered from the beginning of the year January–December 2011 (before the change), and also from Jan 2012 to April 2013 (after the change). The educational awareness activities started in October 2012 and the data was collected on a monthly basis until the end of April. The awareness activities were carried out in different maternity hospitals (ante-natal clinic), women's associations, colleges and universities in Dubai. Moreover, awareness was also carried out using the benefit of the media such as TV programmes, Newspapers and social media. In this project the type of study design is experimental and retrospective-prospective study which focuses on past trends in a phenomenon and studies it into the future. Part of the data is collected retrospectively from the existing records before intervention is introduced, followed by data collected after the impact of the intervention. This kind of design study measures the impact and the effectiveness of awareness activities in altering situations.

6 Results

The most crucial finding from the collected data before and after conducting the educational awareness campaign illustrate the effect of obtaining different awareness activities in promoting knowledge sharing among expectant mothers. This has resulted in an increase in number of collected UCB units both private and public sectors. The start of the awareness activities was in October 2012 to April 2013 in different maternal government hospitals, clinics and women association. Data was collected before and after implanting the HSE change model to be able to analysis the change in number of cord blood collected in period of 6 months. In the UAE, there is no study conducted to test the level of awareness among expectant mothers regarding cord blood stem cells. The author implemented a new approach using HSE model to promote knowledge sharing among expectant mothers in UAE. After the start of the active awareness activities in October 2012–April 2013, the number of collected UCB has increased significantly by 215 units which is 57 % increase by taking into consideration that the total rate of birth in (Mar 2012–Sep2012) was 3483 new birth whereas in (Oct 2012–Apr 2013) the number of birth was even less 3476 birth which demonstrate that the increase in collected UCB during (Oct 2012–Apr 2013) was due to the effective awareness campaign as shown in Fig. 1. The dramatic increase in number of collected units illustrates the effectiveness of change management in promoting knowledge sharing among expectant mothers.

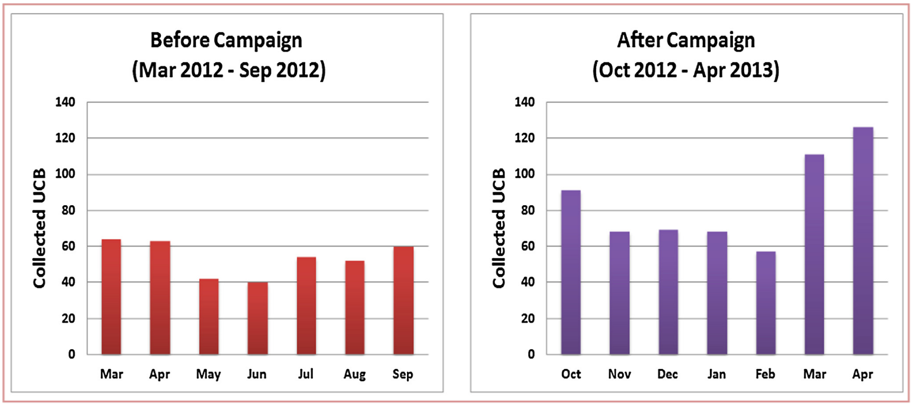


Fig. 1. Affect of awareness campaign on collected UCB units before/after the implementing change.

Figure 1 shows clearly that the collected UCB units on a monthly basis for the period of 7 months before and after the awareness activities. Results show a significant increase in number of collected UCB units after the change by conducting effective educational awareness campaign. The number of privately collected cord blood units has increased significantly by 57 % in just seven months while there was a dramatic increase in number of donated cord blood units overall and this represented almost a 300 % increase.

7 Conclusion

Healthcare organizations should have the culture of knowledge sharing practices to make better use of the knowledge, experiences and skills of their healthcare professionals. This will enable health care professionals to provide better healthcare quality service to the citizens. Knowledge sharing plays a crucial role in promoting knowledge transfer. Our study has shown that through the use of knowledge sharing using a change management approach, it is possible to promote cord blood denotation knowledge to the expectant mothers in Dubai. HSE (Health Service Executive) model was selected to introduce change by conducting different awareness activities about the importance of cord blood in presentations, road-shows, workshops and media events. After the start of the improved educational awareness among expectant mothers the number of collected private and public cord blood units has increased significantly. The finding demonstrates the effectiveness of the successful change management in organization management conducted in promoting awareness activities and knowledge sharing among stakeholders. This study was mainly focusing on promoting knowledge sharing among expectant mothers. However, there are other important stakeholders who are not identified in this paper and play vital role in increased awareness about importance of cord blood stem cells. Further studies should be conducted to identify primary and secondary stakeholders who can help promote knowledge sharing.

Moreover, it is also important to study the best models/methods that help in engaging all relevant stakeholders within the project to ensure stakeholder participation, involvement and knowledge sharing.

References

1. Nilakanta, S., Miller, L., Peer, A., Bojja, V.M.: Contribution of knowledge and knowledge management capability on business processes among healthcare organizations. In: 2nd Hawaii International Conference on System Sciences, HICSS (2009)
2. Orzano, A.J., McInerney, C.R., Scharf, D., Tallia, A.F., Crabtree, B.F.: A knowledge management model: implications for enhancing quality in health care. J. Am. Soc. Inform. Sci. Technol. **59**, 489–505 (2008)
3. Abidi, S.S.R.: Healthcare knowledge management: the art of the possible. In: Riaño, D. (ed.) K4CARE 2007. LNCS (LNAI), vol. 4924, pp. 1–20. Springer, Heidelberg (2008)
4. Health Services Executive (HSE), Improving Our Services – A Users' Guide to Managing Change in the Health Services Executive, Dublin (2008)
5. Wigg, K.: Knowledge Management Foundations. Schema Press, Arlington (1993)
6. Davenport, T., Prusak, L.: Working Knowledge. Harvard Business School Press, Boston (1998)
7. Owayid, A.M., Alrawi, K., Shaalan, K.: Effectiveness of information systems infrastructure and team learning in integrating knowledge management and e- learning technologies. In: The 8th International Conference on Knowledge Management in Organizations, Pp. 557–566, Chapter 45. Springer, Netherlands (2014)
8. Pulakos, E.D., Dorsey, D.W., Borman, W.C.: Hiring for knowledge-based competition. In: Jackson, S.E., Hitt, M.A., Denisi, A.S. (eds.) Managing Knowledge for Sustained Competitive Advantage: Designing Strsteies for Effctive Human Resource Management. John Wiley & Sons, New York (2003)
9. Hooff, V.D.B., Ridder, J.A.: Knowledge sharing in context: the influence of organizational commitment, communication climate and CMC use on knowledge sharing. J. knowl. Manage. **8**(6), 117–130 (2004)
10. Choi, S.Y., Kang, Y.S., Lee, H.: The effects of socio-technical enablers on knowledge sharing: an exploratory examination. J. Inf. Sci. **34**, 742–754 (2008)
11. Lientz, B.P., Kathryn, P.R.: Breakthrough IT Change Management: How to Get Enduring Change Results. Routledge, New York (2004)
12. Burnes, B.: No such thing as a "one best way" to manage organizational change. Manag. Decis. **34**(10), 11–18 (1996)
13. Burnes, B.: Kurt lewin and the planned approach to change: a re-appraisal. J. Manage. Stud. **41**(6), 977–1002 (2004)
14. Liu, Y.: Analysis and evaluation of organizational change approaches. Int. J. Bus. Manage. **4** (12), 234–238 (2009)
15. Bamford, D.R., Forrester, P.L.: Managing planned and emergent change within an operations management environment. Int. J. Oper. Prod. Manage. **23**(5), 546–564 (2003)
16. Sidorko, P.E.: Transforming library and higher education support services: can change models help? Libr. Manage. **29**(4/5), 307–318 (2008)
17. Health Services Executive (HSE), Guiding change in the Irish health system. Dublin (2006)

18. Kotter, J.P., Schlesinger, L.A.: Choosing strategies for change, Harv. Bus. Rev. **57**(2):106–14. https://www.ihm.org.uk/documents/Education/mhsc_resource_file/certificate_manging_people/certificate_managing_people_choosing_strategies_for_change Accessed 18 December 2014 2:00 pm (GMT+4)
19. Brisson-Banks, C.V.: Managing change and transitions: a comparison of different models and their commonalities. Libr. Manage. **31**(4/5), 241–252 (2010)
20. Kotter, J.P.: Leading change: why transformation efforts fail. Harv. Bus. Rev. **85**(1), 96–103 (2007)
21. Alberts, D.J.: A model of multidiscipline teams in knowledge-creating organizations. Team Perform. Manage. **13**(5/6), 172–183 (2007)
22. Smith, B.J., Bauman, A.E., McKenzie, J., Thomas, M.: Awareness of message source and its association with the impacts of sun protection campaigns in Australia. Health Edu. **105**(1), 42–52 (2005)
23. Carlstrom, E.D., Ekman, I.: Organizational culture and change: implementing person-centred care. J. Health Org. Manage. **26**(2), 175–191 (2012)
24. Gulev, R.E.: Cultural repercussions: an analysis of management behaviour through the lens of European cultural variations. Ind. Manage. Data Syst. **109**(6), 793–808 (2009)
25. Agrawal, V.: Managing the diversified team: challenges and strategies for improving performance. Team Perform. Manage.: An Int. J. **18**(7/8), 384–400 (2012)

Improving Hospital Readmission Prediction Using Domain Knowledge Based Virtual Examples

Milan Vukicevic[1(✉)], Sandro Radovanovic[1], Ana Kovacevic[1,2], Gregor Stiglic[3,4], and Zoran Obradovic[2]

[1] Faculty of Organizational Sciences, University of Belgrade, Jove Ilića 154, Belgrade, Serbia
{milan.vukicevic, sandro.radovanovic}@fon.bg.ac.rs, ana.kovacevic@temple.edu

[2] Center for Data Analytics and Biomedical Informathics, Temple University, 1925 North 12th Street, Philadelphia, PA, USA
{zoran.obradovic}@temple.edu

[3] Faculty of Health Sciences, University of Maribor, Zitna Ulica 15, 2000 Maribor, Slovenia

[4] Faculty of Electrical Engineering and Computer Science, University of Maribor, Smetanova 17, 2000 Maribor, Slovenia
gregor.stiglic@um.si

Abstract. In recent years, prediction of 30-day hospital readmission risk received increased interest in the area of Healthcare Predictive Analytics because of high human and financial impact. However, lack of data, high class and feature imbalance, and sparsity of the data make this task so challenging that most of the efforts to produce accurate data-driven readmission predictive models failed. We address these problems by proposing a novel method for generation of virtual examples that exploits synergetic effect of data driven models and domain knowledge by integrating qualitative knowledge and available data as complementary information sources. Domain knowledge, presented in the form of ICD-9 hierarchy of diagnoses, is used to characterize rare or unseen co-morbidities, which presumably have similar outcome according to ICD-9 hierarchy. We evaluate the proposed method on 66,994 pediatric hospital discharge records from California, State Inpatient Databases (SID), Healthcare Cost and Utilization Project (HCUP) in the period from 2009 to 2011, and show improved prediction of 30-day hospital readmission accuracy compared to state-of-the-art alternative methods. We attribute the improvement obtained by the proposed method to the fact that rare diseases have high percentage of readmission, and models based entirely on data usually fail to detect this qualitative information.

Keywords: Virtual examples · Electronic health records · Hospital readmission · Domain knowledge

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 695–706, 2015.
DOI: 10.1007/978-3-319-21009-4_51

1 Introduction

Development of accurate predictive analytics models in healthcare has large benefits for many stakeholders. Hospitals can benefit from healthcare predictive analytics by monitoring of quality indicators, planning of accommodation capacities, optimizing level of supplies, etc. Insurance companies can define adequate charging policies; medical doctors can use decision support in diagnostics, while patients can receive better quality of care, assessment of real costs by different hospitals, etc.

Potential of Electronic Health Records (EHR) predictive analytics applications is recognized, and we are witnessing rapid increase in the number of researchers that are trying to create accurate models for prediction of admission and readmission rates [21], readmission risk [24], cost-to-charge ratio [18], mortality rates, disease prediction [4, 23] etc.

Prediction of 30-day hospital re-admission takes a special place in predictive analytics research. Timely identification of potential un-planned readmissions can have high impact on improvement of healthcare services for patients, by reducing the need for unnecessary interventions and hospital visits. In addition, hospital readmission is considered as one of the most important indicators of quality of care for hospitals, with great economic impact. It is reported that readmission rate within 30 days was 19.6 %, 34.0 % within 90 days, and 56.1 % within one year following discharge. According to the Institute for Healthcare Improvement, of the 5 million U.S. hospital readmissions, approximately 76 % can be prevented, generating annual cost of about $25 billion [2].

Because of its complexity, importance, and involved financial resources (additional costs of several billion dollars annually), this problem is highly regarded in the medical community, and it is frequently addressed in recent research [6]. Unfortunately, researchers often fail to evolve highly accurate predictive models because of high dimensionality, class and feature imbalance, and input space sparsity. Therefore, this problem is highly challenging for the data mining community. Additionally, in this area researchers are often confronted with the lack of data that can show up because of different reasons, e.g. due to long and expensive clinical trials [3] or lack of data due to rare appearance of diseases [1, 11]. In other words, for each prevented re-hospitalization we could save 6,579 USD.

One possible way to address these problems is utilization of virtual examples (VE) [17], used as additional training samples and created from the current set of examples by utilizing specific knowledge about the task at hand. Compared to simple randomization techniques, incorporation of prior knowledge may contain information on a domain not present in the available domain dataset [15] and thus exploits advantages in domain knowledge (knowledge driven) and data driven complementary information sources. Virtual examples are successfully used in computer vision [10, 25] using the knowledge about rotated representation of 3D models and smoothness. In the medical area, for situations where clinical trials are long and expensive, VE that use differential equations [3] or medical models [5] are successfully applied. However most of these models are designed for generation of continuous data and to the best of our knowledge, there is no VE generator that can produce binary data from Electronic Health Records by utilization of domain medical knowledge.

Our contribution: We propose a method for virtual example generation that exploits domain knowledge given in the ICD-9-CM hierarchy of diseases. The proposed technique allows characterization of unobserved comorbidities (disease co-occurrences diagnosed in the moment of patient discharge from hospital) and thus removing bias of algorithms towards frequently observed diseases and comorbidities. We evaluate the proposed method on 66,994 pediatric hospital discharge records from California, State Inpatient Databases (SID), Healthcare Cost and Utilization Project (HCUP) in the period from 2009 to 2011 to illustrate the proposed idea.

2 Related Work

Machine learning algorithms are common tools for prediction of hospital readmission [8, 24]. Steele et al. [22] used a Bayesian Belief Network to predict perioperative risk of clostridium difficile infection following colon surgery. Data used in this research are gathered from the Nationwide Inpatient Sample (NIS) registry from 2005 to 2007. Naive Bayes method is used to select features for the Bayesian Belief Network, which is constructed using data from years 2005 and 2006. Evaluation of the generated model was conducted on 2007 data, having area under curve 0.746. Additionally, authors stated that a lot of data preprocessing is needed in order to get usable results.

Shams et al. [19] created a hybrid classification model which is used to distinguish readmission for heart failure, acute myocardial infarction, pneumonia and chronic obstructive pulmonary disease on 2011-12 Veterans Health Administration data in the State of Michigan. Authors concluded that a lot of efforts were made in order to build risk prediction models, but most of them fail to reach satisfactory accuracy level. A potential reason that authors noticed for this is that developed models often do not make distinctions between planned and unplanned readmissions. Therefore, they developed a new readmission metric that has the ability to identify potentially avoidable readmission from all other types of readmission. Results obtained on the proposed readmission metric were very promising, with area under curve just over 0.8 for all above-mentioned diseases.

Even though many efforts are invested in EHR analysis and predictive modeling, and secondary use of EHR has large potential, researchers are not satisfied with the results for diagnosis or readmission prediction. Therefore, Ooi et al. [16] advise that machine learning alone cannot be used for this purpose. They propose a hybrid human-machine database engine where the machine interacts with the subject matter experts as part of a feedback loop to gather, infer, ascertain, and enhance the database knowledge and processing and discuss the challenges towards building such a system through examples in healthcare predictive analysis. McCoy et al. [14] also stress that machine learning on only EHR records is not sufficient and that clinical text is a major component of EHR data and often contains rich information describing patients.

However, in situations with the lack of data, high class or feature imbalance machine learning algorithms have limited potential. In previous studies, for addressing such problems, virtual examples have been mainly used for dealing with limited observations or low quality data [85]. Several studies proposed powerful tools for effective generation of virtual examples based on specific prior knowledge in domain specific applications. For example, in computer vision, rotated representation of 3D models and smoothness is successfully used to consider the integrated effects and constraints of data attributes by using mega-trend diffusion functions, and feasibility-based programming models [10, 25]. They showed that oversampling with VE improves performance and stability of learning algorithms.

In the area of healthcare predictive modeling, virtual examples are successfully used for sepsis analysis. VE are of crucial significance for early sepsis prediction, since patients infected by this disease often die in the early stage and thus temporal data cannot be gathered. Recently proposed predictive models for addressing this problem are based on VE that use differential equations [3], or medical models [5] as prior knowledge sources.

It can be concluded that VE are a useful, and sometimes the only possible, way to address the problem of lack (or imbalance) of data in predictive modeling. To the best of our knowledge there is no VE generator that uses domain knowledge in order to enrich EHR. Therefore, we developed a virtual examples generator that uses domain knowledge information extracted from ICD-9 hierarchy of diseases and complement HER records about 30-day hospital re-admission prediction.

3 ICD-9 International Classification of Diseases

The ICD-9 EHR codes are organized in a hierarchy where an edge represents 'is-a' relationship between a parent and its children. Hence, the codes become more specific as we go down the hierarchy [20]. Each three digit diagnostic code (group of diagnoses) is associated with a hierarchy tree of ICD codes on lower level. In this paper, we refer to it as a 'top-level' diagnostic code. Figure 1 shows a part of the hierarchy within the top-level (most general) diagnostic code 530 that represents diseases of the esophagus. Top-level can be represented as a set of four level diagnoses, which present more specific diagnoses. Further, one four digit code can be specified to more specific concepts (five digit codes).

4 Exploratory Analysis of Pediatric 30-Day Hospital Readmission Data

In this research, hospital discharge data from California, State Inpatient Databases (SID), Healthcare Cost and Utilization Project (HCUP) [7], Agency for Healthcare Research and Quality was used. This data tracks all hospital admissions at the individual level, having a maximum of 25 diagnoses for each admission. Every diagnosis is presented as an ICD-9-CM code. Beside diagnoses, every admission is explained with an additional 21 features (e.g., sex, age, month of admission, length of stay, total charges in USD, etc.). We used all pediatric patient data in California from January 2009 through December 2011 in the pre-processing phase.

Since there are over 14,000 ICD-9-CM codes, usage of codes as categorical features would be infeasible. Therefore, we transformed diagnoses to binary features, where if a patient had that diagnosis, value would be positive, otherwise negative.

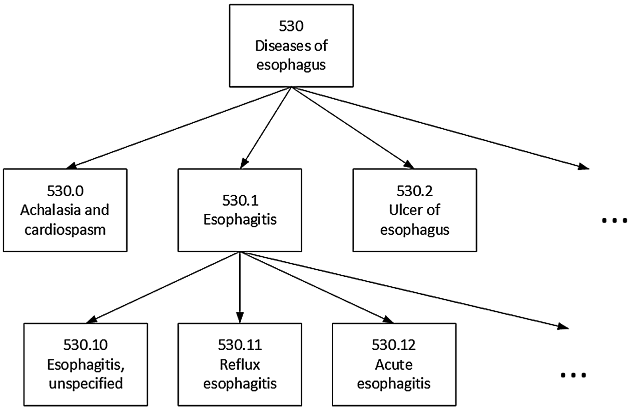


Fig. 1. Excerpt of three digit ICD-9 hierarchy

With this step, we transformed 15 features with very high cardinality (over 14,000 possible values) to over 14,000 binary features. Further, we excluded features with zero positive values. After this transformation, 851 ICD-9-CM codes remained, leading to a total of 872 binary features that we used to predict whether a patient will be readmitted within 30 days.

In order to characterize examples that could potentially be useful for improvement of the accuracy of predictive algorithms, we conducted exploratory analyses of the data with regard to the knowledge provided by ICD-9 hierarchy.

First, we investigated the balance between appearance of different diseases (on the lowest level of hierarchy) and their contribution to cumulative re-admission. Figure 2 (left), shows that most of the diseases appear rarely. In contrast, in cumulative, rare diseases constitute a large portion of re-admissions. This can be clearly seen on Fig. 2 (right), where on the X-axis diseases are represented in ascending order by the frequency of appearance, while on the Y-axis, cumulative share of each disease in total number of readmissions is showed. In particular, point (2000, 0.8) in this figure means that 80 % of readmissions are related to 1000/14000 (or, 14 %) of the least common primary admission diagnoses.

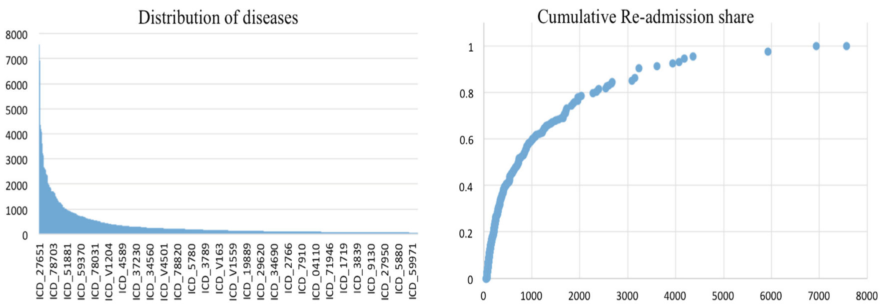


Fig. 2. Distribution of diseases (left) and cumulative readmission rate (right)

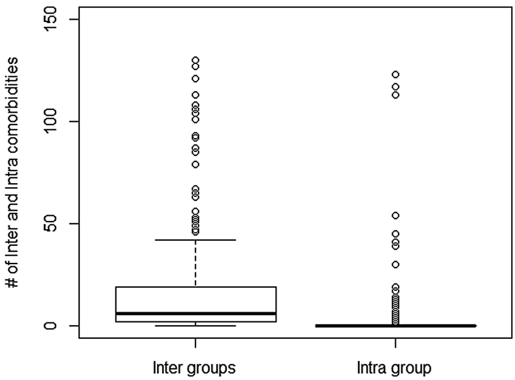


Fig. 3. Comparison of inter and intra 3digit ICD groups comorbidity frequencies

The large share of total re-admissions that is caused by rare diseases induces a problem for evolving of accurate predictive algorithms, since most of them are biased towards frequent features. Further, we expanded our research on investigation of comorbidity (co-occurrences of different diseases in each patient record) frequencies and found that comorbidities between ICD9-3 digit groups are much more frequent than comorbidities within groups. This is clearly shown on Box Plots in Fig. 3.

Based on previous observation, we focused on the inter-group comorbidities and further investigated their relation with re-admission. We analyzed standard deviation of re-admission rate between comorbidities from one group with diseases in other groups. The main hypothesis here was that if diseases from one ICD9-3digit group have small standard deviation of re-admission rate when they appear as comorbidities with diseases from other groups, then we can generate virtual examples that will connect rare (or even un-observed) diseases with high confidence about their re-admission. Analyses of standard deviation of readmission risk by all 3-digit groups (Fig. 4) showed that each 3-digit group has average standard deviation of re-admission less than 0.2. Most of them are below 0.11 or even smaller.

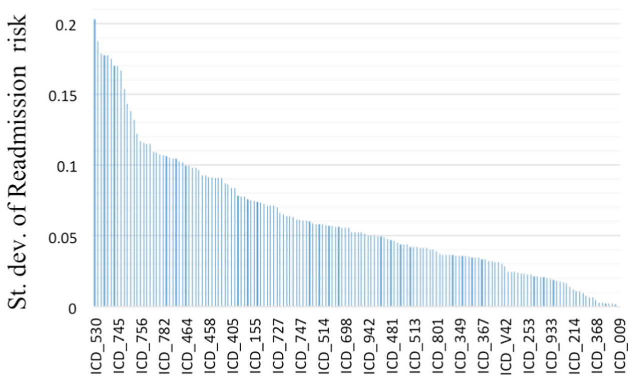


Fig. 4. Average standard deviation of re-admission between diseases of one group with other groups

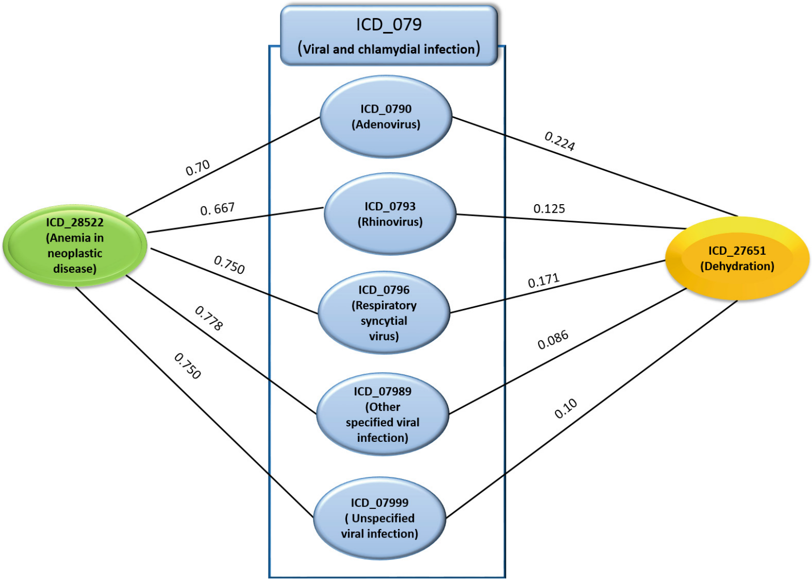


Fig. 5. Similar readmission risk of comorbidities from 3-digit group

This finding allowed generation of virtual examples that could potentially improve accuracy of predictive algorithms. In the following section we propose a data generator that implements findings from exploratory analyses of 30-day Hospital re-admission data. Figure 5, shows the situation where all diseases from one 3-digit group (central part) have similar readmission risks (RR) when connected with other diseases (all diseases from ICD_079 have high RR when co-occur with ICD_28522 and low RR when co-occur with ICD_27651).

5 ICD-9-VEG: Knowledge Based VE Generator

Based on considerations from the previous section we propose an ICD-9 based virtual example generator (ICD9-VEG). In order to address the problem of imbalanced features (since it is observed that rare diseases make large portion of re-admissions) generation of new examples (patient discharge record) goes as follows: First *a priory* probability of disease appearance is calculated for each disease. Initial disease is selected based on inverted probabilities (largest probabilities for selection have rare diseases).

When the first diagnosis is selected, comorbidity subset (CS) is formed from all diagnoses that have comorbidities with the selected diagnosis. The next feature is chosen based on λ-updated probabilities from CS with probability update not only from comorbidities with previously selected diseases, but all comorbidities of 3-digit hierarchy level that selected diagnoses originate from. This extends the space of possible diagnosis and allows knowledge-guided selection of unseen cases. Intuition behind this approach is that diagnoses from the same hierarchical group are often treated the same way, having similar symptoms and diagnostic criteria (i.e. [9]), meaning that real diagnosis could be overlooked. This intuition is also confirmed in Sect. 4. Further, after selection of each disease, the decision about adding more comorbidities is made based on conditional probability that selected subset of diseases appears with other diseases. Thus, if examples in CS have more diagnoses than new virtual examples, then probability of continuing is greater. In order to enable generation of unseen comorbidities we set up a parameter of continuation that allows adding new comorbidities even if they are not observed in the initial set. Finally, when all diseases are created for the record, readmission outcome (true or false) is assigned based on readmission risk of the CS selected.

When new virtual examples are created they are stored in the VE variable, but they are also added to the initial dataset. The reason for this step is explained in Sect. 3, where we observed that "rarely" observed diseases and comorbidities make large portion of re-admissions. With this step we perform probability averaging of diagnoses, which leads to balancing of disease appearance in the initial data. Level of randomization and ICD9 influence is controlled by a smoothing parameter that controls smoothing level (by increasing **λ,** probability of generating unseen comorbidities grows). Users also provide the number of examples to be generated a parameter for smoothing variables other than diagnosis. The flow of the ICD9-VEG algorithm is depicted on Fig. 6.

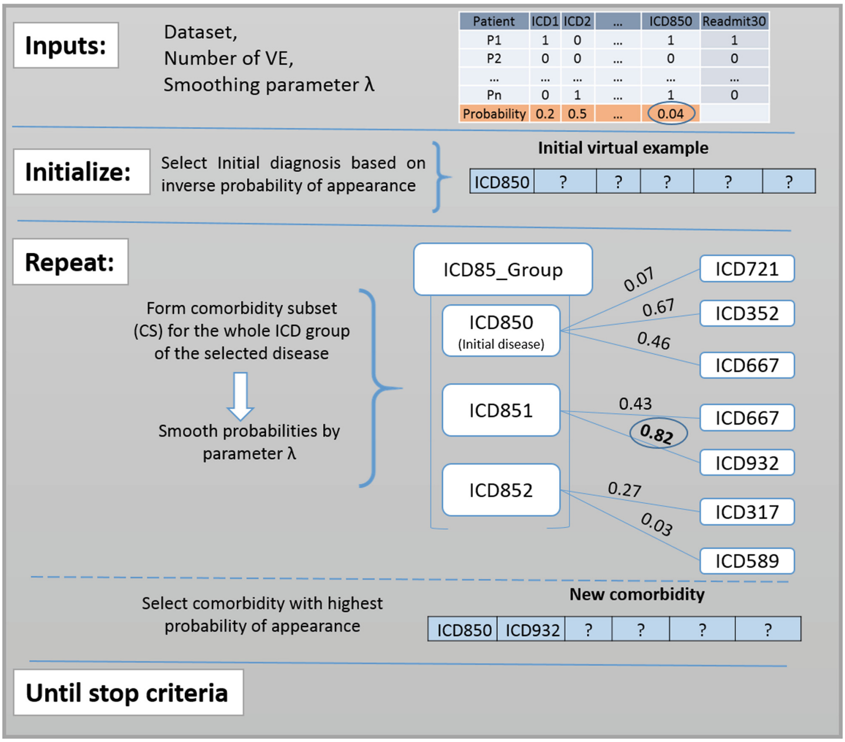


Fig. 6. Description of ICD9-VEG generator

6 Experimental Evaluation

In order to evaluate an influence of virtual examples on the predictive performance of algorithms we used Logistic regression and Naïve Bayes, because of their interpretability and success in healthcare applications [24]. Evaluation is performed on all pediatric patient data (described in previous section) where patient records that were admitted in period 2009-2010 were used for training (46,682 cases), and 2011 for testing (20,312 cases).

Influence of adding virtual examples generated by ICD9-VEG is compared with commonly used oversampling. We created two oversampling sets as a benchmark. First, "Class balance" was prepared by oversampling minority class to completely balance classes (37872 examples with positive re-admission rate are replicated from the initial 7611 samples, and added to the training set). In order to inspect the influence of using domain knowledge in the data generation process we prepared another benchmark set by using similar strategy as ICD9-VEG. "Feature balance" set is prepared by oversampling records that have rare diagnosis (we used the same technique of randomization as in ICD-9VEG). This way the influence of incorporating domain knowledge can be clearly explained and fair comparison of the approaches is made possible.

Additionally, we compared our approach to common ensemble algorithms such as AdaBoost and Bagging. AdaBoost and Bagging are performed in 10 iterations, where Bagging selected 70 % of the dataset with replacement. Experiments are performed 10 times and average is presented. Under-sampling is not considered in this research, since a huge amount of information would be lost.

Since hospital re-admission is highly class imbalanced (in our case there are 11,884 positive and 55,810 negative cases in all three years), in order to inspect in more detail the influence of new information on predictive performance we used four accuracy measures: Area Under Curve (AUC), Precision, Recall and F1 score. Table 1 shows performance of Logistic regression with and without oversampling or boosting techniques.

It can be seen from Table 1 that ICD9-VEG based oversampling gave the best results measured by AUC, precision, and F1 score. It is interesting to note that the "Class balance" technique led to the best recall. This means that Logistic regression managed to predict accurately the largest percentage of readmitted patients. On the

Table 1. Logistic regression performance

Logistic Regression	AUC	Precision	Recall	F1 score
ICD9-VEG	**0.802**	**0.715**	0.357	**0.476**
Logistic regression	0.787	0.693	0.344	0.460
Feature balance	0.775	0.670	0.313	0.427
Class balance	0.784	0.086	**0.366**	0.139
Bagging	0.776	0.669	0.350	0.460
Boosting	0.740	0.693	0.344	0.460

Table 2. Naïve Bayes performance

Naïve Bayes	AUC	Precision	Recall	F1 score
ICD9-VEG	**0.745**	**0.412**	0.683	**0.472**
Naïve Bayes	0.668	0.341	0.664	0.451
Feature balance	0.672	0.339	0.677	0.455
Class balance	0.659	0.080	0.308	0.127
Bagging	0.718	0.326	**0.706**	0.446
Boosting	0.701	0.371	0.664	0.451

other hand, Precision by Class balance is drastically lower compared to other techniques. This means that this type of oversampling biased the classifier to the minority class, and thus it almost always predicts that class. This is reflected in AUC and F1 as general measures of quality of classifiers. In Table 2, the results for Naïve Bayes are showed. It can be seen that performance of Naïve Bayes was worse than Logistic regression and showed higher tendency of predicting rare class (low precision and high recall). The pattern of performance by different oversampling and ensemble techniques stayed the same as for Logistic regression: ICD9-VEG was again the best performing technique by all measures except recall, where Bagging takes the lead.

To take the results into perspective we can estimate the financial impact of our results according to a study by Behara et al. [2] where an estimated cost of prevented re-admission accounts to 6,579 USD. Observing precision, out of 1000 positively classified selected patients, our approach will correctly identify 412 of them as positive. In comparison to Naïve Bayes this is 71 patients more which amounts to almost 0.5 million USD in savings for only 1000 patients selected by both methods (assuming that we can treat each patient in a way that will prevent re-admission).

Discussion: We showed that incorporating domain knowledge from ICD-9 hierarchy into oversampling techniques can improve classifier performance. We attribute that to the fact that rare diseases are the ones that have high percentage of readmission, and models usually fail to detect this information. By generating VEs which use this information, we obtain better performance. It is worth noticing that AdaBoost and Bagging didn't improved performance for logistic regression, which indicates over-fitting. On the other hand, the ensemble approach did improve performance in the Naïve Bayes algorithm. For both Logistic Regression and Naïve Bayes, class balancing with oversampling showed biased results towards minority class (positive re-admission) and feature balancing without incorporation of prior knowledge slightly reduced the performance.

7 Conclusion and Future Research

In this study we addressed the problem of insufficient training data and feature imbalance in healthcare predictive analytics by proposing a method for virtual example generation that exploits domain knowledge and available data as complementary sources. Experimental evaluation showed that this strategy improves predictive performance of popular classification algorithms and shows competitive performance with traditional oversampling and ensemble strategies. Additionally, the proposed technique allows generation of unobserved comorbidities and thus removes bias of algorithms towards frequently observed diseases and comorbidities. The proposed approach is therefore useful in cases of rare conditions with low prevalence or even in cases where we would like to observe the predictive performance of classification models in unseen scenarios (e.g. disease outbreaks, epidemics or pandemics). As a part of our future work, we will extend the proposed framework by different sources of medical prior knowledge e.g. patient, disease, drug ontology. Additionally, we plan to extend the virtual example generator to generation of continuous data. This will allow generation of more input/output features that are available from original data (i.e. patients age or length of stay in the hospital). The extended generator will be used for enriching EHR repositories for improving predicting performance of the algorithms on readmission prediction as well as other healthcare related tasks, such as length of stay, admission, and mortality rate predictions.

Acknowledgments. This research was supported by DARPA Grant FA9550-12-1-0406 negotiated by AFOSR, National Science Foundation through major research instrumentation, grant number CNS-09-58854, and by SNSF Joint Research project (SCOPES), ID: IZ73Z0_152415.

References

1. Bavisetty, S., Grody, W.W., Yazdani, S.: Emergence of pediatric rare diseases: review of present policies and opportunities for improvement. Rare Dis. **1**(1) (2013)
2. Behara, R., Agarwal, A., Fatteh, F., Furht, B.: Predicting Hospital Readmission Risk for COPD Using EHR Information. Handbook of Medical and Healthcare Technologies, pp. 297–308. Springer, New York (2013)
3. Cao, X.H., Stojkovic, I., Obradovic, Z.: Predicting sepsis severity from limited temporal observations. In: Džeroski, S., Panov, P., Kocev, D., Todorovski, L. (eds.) DS 2014. LNCS, vol. 8777, pp. 37–48. Springer, Heidelberg (2014)
4. Davis, D.A., Chawla, N.V., Christakis, N.A., Barabási, A.L.: Time to CARE: a collaborative engine for practical disease prediction. Data Min. Knowl. Disc. **20**(3), 388–415 (2010)
5. Ghalwash, M., Obradovic, Z.: A data-driven model for optimizing therapy duration for septic patients. In: Proceedings 14th SIAM Int'l Conference Data Mining, 3rd Workshop on Data Mining for Medicine and Healthcare, Philadelphia, April 2014
6. Hasan, O., Meltzer, D.O., Shaykevich, S.A., Bell, C.M., et al.: Hospital readmission in general medicine patients: a prediction model. J. Gen. Intern. Med. **25**(3), 211–219 (2010)
7. HCUP State Inpatient Databases (SID), Healthcare Cost and Utilization Project (HCUP). 2009–2011. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. 27 Feb 2015
8. Krumholz, H.M., Lin, Z., Drye, E.E., Desai, M.M., Han, L.F., Rapp, M.T., Normand, S.L.T.: An administrative claims measure suitable for profiling hospital performance based on 30-day all-cause readmission rates among patients with acute myocardial infarction. Cir. Cardiovas. Qual. Outcomes **4**(2), 243–252 (2011)

9. Legname, G.: Novel approaches to diagnosis and therapy in neurodegenerative diseases. In: Toi, V.V. (ed.). IFMBE Proceedings, vol. 46, pp. 155–158Springer, Heidelberg (2015)
10. Li, D.C., Wen, I.H.: A genetic algorithm-based virtual sample generation technique to improve small data set learning. Neurocomputing **143**, 222–230 (2014)
11. Mathew, G., Obradovic, Z.: Distributed privacy preserving decision system for predicting hospitalization risks in hospitals with insufficient data. In: Machine Learning in Health Informatics Workshop: International Conference on Machine Learning Applications–ICMLA, pp. 178–183. Boca Raton, FL, USA (2012)
12. Mathew, G., Obradovic, Z.: Distributed privacy preserving decision support system for highly imbalanced clinical data. ACM Transactions on Management Information Systems. **4** (3) Article No. 12, October 2013 (2013)
13. Mathew, G., Obradovic, Z.: A distributed decision support algorithm that preserves personal privacy. J. Intell. Inf. Syst. **44**, 107–132 (2014)
14. McCoy, A.B., Wright, A., Eysenbach, G., Malin, B.A., Patterson, E.S., Xu, H., Sittig, D.F.: State of the art in clinical informatics: evidence and examples. Yearb. Med. Inform. **8**, 13–19 (2013)
15. Mirchevska, V., Luštrek, M., Gams, M.: Combining domain knowledge and machine learning for robust fall detection. Expert Syst. **31**(2), 163–175 (2014)
16. Ooi, B.C., Tan, K.L., Tran, Q.T., Yip, J.W., Chen, G., Ling, Z.J., Zhang, M.: Contextual crowd intelligence. ACM SIGKDD Explor. Newsl. **16**(1), 39–46 (2014)
17. Poggio, T., Vetter, T.: Recognition and structure from one 2D model view: Ob- servations on prototypes, object classes and symmetries. Artificial Intell. Lab., MIT, Cambridge, MA, A.I. Memo no. 1347, April 1992 (1992)
18. Polychronopoulou, A., Obradovic, Z.: Hospital pricing estimation by gaussian conditional random fields based regression on graphs. In: Proceedings of the 2014 IEEE International Conference on Bioinformatics and Biomedicine, Belfast, UK, Nov 2014
19. Shams, I., Ajorlou, S., Yang, K.: A predictive analytics approach to reducing 30-day avoidable readmissions among patients with heart failure, acute myocardial infarction, pneumonia, or COPD. Health care management science, 1–16 (2014)
20. Singh, A., Nadkarni, G., Guttag, J., Bottinger, E.: Leveraging hierarchy in medical codes for predictive modeling. In: Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics, pp. 96–103. ACM, September 2014
21. Srivastava, R., Keren, R.: Pediatric readmissions as a hospital quality measure. JAMA **309** (4), 396–398 (2013)
22. Steele, S., Bilchik, A., Eberhardt, J., Kalina, P., Nissan, A., Johnson, E., Stojadinovic, A.: Using machine-learned bayesian belief networks to predict perioperative risk of clostridium difficile infection following colon surgery. Interactive J. Med. Res. **1**(2) (2012)
23. Stiglic, G., Pernek, I., Kokol, P., Obradovic, Z.: Disease prediction based on prior knowledge. In: ACM SIGKDD Workshop on Health Informatics, in Conjunction with the 18th SIGKDD Conference Knowledge Discovery and Data Mining, Beijing, China, Aug 2012
24. Stiglic, G., Wang, F., Davey, A., Obradovic, Z.: Readmission classification using stacked regularized logistic regression models. In: Proceedings of the AMIA 2014 Annual Symposium, Washington, DC, Nov 2014
25. Yang, J., Yu, X., Xie, Z.Q., Zhang, J.P.: A novel virtual sample generation method based on Gaussian distribution. Knowl. Based Syst. **24**(6), 740–748 (2011)

Safety Use of Hospital Information Systems: A Preliminary Investigation

Lizawati Salahuddin(✉) and Zuraini Ismail

Advanced Informatics School (AIS), Universiti Teknologi Malaysia (UTM),
54100 Kuala Lumpur, Malaysia
lizawati22@live.utm.my, zurainiismail.kl@utm.my

Abstract. One of the strategies to enhance patient safety due to medical errors is through the use of Hospital Information Systems (HIS). However, previous studies have revealed that its usage introduces new class of errors called technology-induced errors. These errors occurred when HIS is poorly designed, implemented or used. The antecedents of HIS related errors originated from various factors. It is essential to ascertain the root cause for safe use of HIS in order to reduce the technology-induced errors. Hence, this preliminary study is aimed to explore and identify the antecedents towards safety use of HIS. Semi-structured interviews were conducted at a teaching university hospital, a government hospital, and the Malaysian Ministry of Health from May till July 2014. Six respondents who are experienced HIS users, and also involved in the development of HIS were individually interviewed. Six themes emerged from the semi-structured interviews. This study adds knowledge to HIS research by considering sociotechnical aspect of safety HIS usage.

Keywords: Hospital information systems · Patient safety · Technology-induced errors · Sociotechnical

1 Introduction

Hospital information systems (HIS) is a computerized integrated information system that manages hospitals' administrative, financial and medical information [1]. HIS is developed to support healthcare organization in providing efficient, quality and safe healthcare services. HIS may lead to a safer care by improving communication among healthcare practitioners, and facilitating shared decision making [2]. Nevertheless, HIS involves interaction between healthcare practitioner and complex sociotechnical system. Thus, HIS may introduce new safety risks such as dosage errors, delay in detection of fatal illnesses, and delayed treatment [2]. The safety risks can lead to safety incidents which could have resulted, or did result, in unnecessary harm to a patient [3].

The use of HIS resulted in new types of errors namely technology-induced errors. These errors are influenced by numerous factors. It may attained during the design and development of a technology, the implementation and customization of a technology, and the adoption of a technology [4]. This may lead to deferment from the aims of HIS usage.

Despite tagged as a new developed country, Malaysian government has achieved significant progress in the development and utilization of information technology

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 707–721, 2015.
DOI: 10.1007/978-3-319-21009-4_52

(IT) in healthcare sector. Under the Tele-health Flagship Applications, the Malaysian Ministry of Health (MOH) has embarked on the introduction of electronic HIS in several public hospitals under the Total Hospital Information System (THIS) project [5]. The objective of the project is to establish a paperless hospital environment through a comprehensive information communication and technology (ICT) system, subsequently offering quality health services. Although THIS has been implemented, medical errors particularly medication errors had resulted patient harm [6]. There is paucity on studies related to HIS safety in Malaysia context. To date only one study on HIS safety has been conducted in Malaysia setting [7]. The study conducted interviews with medical practitioners, pharmacists, staff nurse and IT staffs in three Malaysian private hospitals. The safety evaluation of clinical information system in the study was merely looking into the technology perspective. Thus, it is crucial to identify the antecedents towards the safety use of HIS in Malaysia context.

This preliminary study aims to explore and identify the antecedents towards safety use of HIS especially in Malaysia context through qualitative methodology. This paper is organized into seven sections. This section introduced HIS usage. The following second section provides the research background. Next, the third section describes the research methodology and followed with data analysis in the fourth section. Subsequently, the fifth and sixth section elaborates the findings from the semi-structured interviews and discussion respectively. The last section concludes and suggests future works.

2 Research Background

Healthcare system is a complex and high-risk system [8]. In critical care, the complexity of performing tasks is augmented by the constrictions of time, inadequate or unavailable information, by stress, and by repeated and unpredictable interruptions [9]. The tasks are frequently context-dependent, unpredictable, interrupted, and depend on coherent and timely communication between different healthcare practitioners [10]. Therefore, the interdependent nature facilitate the propagation of errors such that any error created by one component may affect other components as well which is normally unpredictable [8]. Accordingly, Hospital Information Systems (HIS) has been introduced to improve the quality of care and patient safety [11].

HIS can prevent errors and adverse events [12]. Nevertheless, there are evidence indicating that HIS can cause patient harm, injury, disability and death [13, 14]. The US Food and Drug Administration (FDA) reported 42 reports of patient harm and four deaths in 436 critical incidents involving health information technology (IT) over a 30-month period, from January 2008 to July 2010 [13]. A recent study analyzed 100 unique and closed investigations between August 2009 and May 2013 from 344 reported incidents revealed 74 of the safety concerns involved unsafe technology, whereas 25 involved unsafe use of technology [15]. In another study explored types of e-prescribing errors and their potential consequences in five community pharmacies in US [16]. They found that 75 e-prescribing errors were documented during 45 h of observation. The factors that contribute to the errors include technology incompatibility between pharmacy and clinic systems, technology design issues, and inadvertently entering incorrect information. In general, the studies showed that the antecedents towards safety use of HIS are not solely dependent upon the technology but also influenced by sociotechnical aspects.

Current preventing actions are largely focusing on the improvement of software design [17, 18]. To date, most of the current standards and guidelines emphasized on the safe design and implementation of software. Certainly, it is essential to understand antecedents towards safe use of HIS particularly from the perspective of sociotechnical approach. As such, this can provide further insights into safety implementation of HIS as a tool to improve quality of healthcare and patient safety.

3 Research Methodology

While waiting for ethics approval from the Malaysian Ministry of Health (MOH), a preliminary investigation was carried out. Therefore, the interviewees were selected based on voluntary basis. The interview study was conducted according to procedural steps suggested by [19] as follows:

1. Based on the literature review findings, interview questions were developed.
2. Potential respondents were selected based on their experiences in using HIS or were involved in the development of HIS. The inclusion criteria is the time duration, by which a minimum two years experience in using HIS or involved in HIS development was outlined.
3. The selected respondents were contacted by the researcher through e-mail and telephone calls. Upon agreement to participate in the study, an interview appointment was set based on their preferences.
4. Semi-structured interviews were conducted with the respondents.

Convenience sampling was adopted in selecting the interviewees. The drawback of convenience sampling is it limits the ability to generalize the results. However, the objective of this research is not generalization, but to explore and identify the antecedents towards safety use of HIS. Only six people from three organizations namely Malaysian government hospital, a teaching university hospital, and MOH agreed to be interviewed. The interviewees were two hospital doctors and pharmacists respectively, a HIS developer, and a deputy director in Malaysian MOH who was involved in the implementation of HIS in Malaysian government hospitals. The semi-structured interviews were conducted between May 2014 and July 2014. The respondents' details are summarized in Table 1. The interviewees were all experienced in their fields, with a minimum of 8 years involvement in the HIS.

The interviews were audio-taped and were transcribed verbatim. All interviews were completed between 30 min and 1 h 20 min. The semi-structured interviews were guided by the research questions which were:

1. What is your opinion on errors resulted from HIS usage?
2. What are the antecedents that influence HIS related errors?
3. What is the necessary safety procedure in ensuring the safety of HIS usage?

Table 1. Respondents' details for preliminary investigation

No.	Designation	Specialization	Organization	Working Experiences
Respondent 1	Doctor	HIS user	Hospital	12 years
Respondent 2	Doctor	HIS user	Hospital	8 years
Respondent 3	Pharmacist	HIS user	Hospital	8 years
Respondent 4	Pharmacist	HIS user	Hospital	9 years
Respondent 5	Deputy Director	HIS development	MOH	15 years
Respondent 6	IT Officer	HIS development	Hospital	8 years

4 Data Analysis

After the semi-structured interview, data analysis was carried out. The analysis involved four phases namely transcribing, organizing, coding, and themes building. First, audio recorded data and handwritten notes gained from interviews were transcribed to word processor text. Then, the interview transcripts will be organized into sections for easy retrieval in the organizing phase. Subsequently, the transcribe interviews were coded. In the coding phase, the transcripts were read repeatedly to highlight parts of the text, and emphasize the sections and issues that seemed important and relevant. The interviews transcripts were divided into text segments, and these segments were labeled with codes. Any overlapping and redundant codes were eliminated. Finally, similar codes were grouped together as a theme or category to form a major idea in the themes building phase [20].

5 Findings

Six themes were identified from the interviews. They are (1) the use of HIS resulted in errors, (2) system quality, (3) information quality, (4) training, (5) task-related stressor, and (6) HIS safety use procedure. The following sections further explain the each of the theme.

5.1 The Use of HIS Resulted in Errors

Majority of the respondents expressed their concerns about the use of HIS resulted in errors. They reported that the interaction between HIS and human can result in errors. The examples of errors related to the HIS usage include wrong patient, medication, dose, and x-ray result. Three respondents conveyed their views stating:

> "Wrong patient, medication, dose, x-ray result are the common errors during the use of HIS". (Respondent 2)
> "HIS can caused errors. Wrong dose and medication are the example of HIS usage errors happened here." (Respondent 3)
> "Errors can happen. For an example, a nurse wrongly selected patient name which appears similar. It did happen to me. Luckily I realized that when I discovered that the name on the prescription was not my family's name. It was someone else's." (Responded 6)

Through the interview it was discovered that human errors were the frequent contributors to the errors in the use of HIS. Wrong data entry, the healthcare delivery procedure related information generated by HIS was overlooked, and data documentation errors are among the examples of the HIS usage resulted to errors. Two respondents described the following views. They said:

> "Errors in the usage of HIS are due to wrong data entry or overlook on the healthcare delivery related information generated by HIS." (Respondent 2)
> "Documentation is done by human, thus prone to error." (Respondent 5)

One respondent emphasized that the HIS system and manual procedure must follow the same order of healthcare delivery procedures in order to improve patient safety through the use of HIS:

> "No matter how well the system is, error still can happen because the system is operated by human. System and human must be aligned so that HIS can improve patient safety." (Respondent 4)

Based on the interview findings, selecting wrong information such as patient name, medication, dose, and medical results were the common errors resulted from the usage of HIS. Human factor seems to suggest a typical reason for errors. Therefore, the implementation of HIS must emphasize on the sociotechnical perspective in order to maximize the potential benefits for the use HIS while minimizing errors.

5.2 System Quality

System quality are concerned with the system features of HIS [21]. System quality is typically measured in terms of adaptability, availability, reliability, response time, usability, and compatibility [21, 22]. From the interviews, it was found that system quality issues were associated with system usability, and compatibility.

ISO 9241-210 standard defines usability as the extent to which a system can be used by specific users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use. The interview findings showed that usability is associated with the ease of use, and illogical ordering of the medication terms in the HIS interface. Ease of use is related to the design of the user interface that enables the user to use the system efficiently. Ease of use is crucial to ensure the smooth of workflow. One respondent remarked that:

> "The system is not user friendly. Every step in the HIS requires password which is frequently time consuming and interrupt the efficiency of our workflow progress." (Respondent 4)
> "We can only open one screen function at one time. The system does not allow us to open multiple screens that represent different functions screen simultaneously, and thus inhibit us from viewing other related information easily." (Respondent 4)

Respondents also indicated that illogical ordering of the terms in the HIS interface for instance medication name may lead to errors such as wrong drug selection. The medication terms should display in logical order according to the healthcare practitioners' perspective as highlighted by one respondent.

> "The drug name is listed in the alphabetical order rather than common drugs taken by patients. This arrangement may cause the doctors to easily click on the wrong drug which are adjacent to the intended drug listed." (Respondent 4)

Appropriate display of medical related information in the HIS interface is important. The healthcare practitioners are alerted on the important information in order to ensure they do not miss any information. One respondent highlighted this view as:

> "User interface must appropriately highlight abnormal laboratory result such as using blinking font or by striking color such as red to attract the users' attention." (Respondent 1)

System compatibility refers to the fit, alignment or balance that needs to be in place to achieve one or more goals [23]. Systems should provide sufficient functions to support the users' tasks. HIS lacks in alert functions to prevent medical errors especially during data entry. This finding is illustrated by comments of one respondent who said:

> "Not all of the required functions are available in the system. We have to explore the system functions by ourselves in order to adapt to the task we need to execute." (Respondent 4)
> "Fortunately, the pharmacist detected wrong dosage being ordered through the system. A doctor wrongly typed 4000 mg instead of 400 mg. This happened due to no safety alert was embedded within the system. In another case, a junior doctor forgot to change the default dosage which was meant for a child. The default dosage was set according to the adults' dosage which is 600 mg." (Respondent 4)

Moreover, systems should be well integrated with other systems as healthcare involves various systems and machines. Improper system integration potentially poses risk to patient safety. A respondent highlighted his view as:

> "Healthcare involves complex and various machines. Integration between different systems and between system and machine are not straightforward. It must comply with variety of the system and machine standards. Inappropriate system integration can jeopardize patient safety." (Respondent 1)

Hence, the findings highlighted the importance of system quality in terms of usability and compatibility may influence the safety use of HIS. Poor usability hinders the healthcare practitioners to use the HIS effectively. Poor system compatibility for instance lack of necessary functions provided by HIS, and integration problem with other related systems or devices possibly pose risk to patient safety.

5.3 Information Quality

Information quality is concerned with the information provided by HIS. Information quality can be measured in terms of accuracy, timeliness, completeness, relevance, and consistency [21]. The interview findings showed that information quality related to the safety issue is denoted by information relevancy. Information relevancy is defined as the ability of the HIS generates information or features that satisfy with the healthcare practitioners' needs [24]. Irrelevance alert generated by HIS resulted annoyance, and consequently caused healthcare practitioners to ignore to the alerts. This leads to failure to act to a truly important warning that may cause danger to patients' safety. One respondent voiced his view as:

> "Alert is one of the features embedded into HIS in addressing the patient safety. However, alert fatigue and out of range alerts caused some of the users to ignore the alert." (Respondent 5)

From the interview outcomes, information quality is another antecedent associated with HIS that may influence the safety use of HIS. Irrelevance information such as alert fatigue or excessive of alert generated by the HIS caused some of the healthcare practitioners neglected the alerts. As the result, the healthcare practitioners may not responded appropriately especially when the warning concerns about patient safety.

5.4 Training

It was revealed that training influences the use of HIS. Training provides knowledge to enable the healthcare practitioners to use the HIS appropriately. Inadequate training cause healthcare practitioners to spend more time in using HIS. One respondent reported his view as:

> "Lack of training caused the healthcare practitioners be incompetent in using the system and at the same time slows down the healthcare process." (Respondent 1)

Training should also provide knowledge on the safety aspect of the HIS usage. One respondent remarked that:

> "Training on system influences the healthcare practitioner to use the HIS correctly. However, system development was outsourced from foreign vendor in which the necessary patient safety aspects were taken into considerations. But, I do not know whether the training emphasized on the safety or not." (Respondent 5)

The training was conducted by vendor under the purchasing contract. More training are required outside the scheduled set by vendor. This results in additional expenses is not within organization's budget. The number of the required training is dependent on the number of new staff recruitment. The recruitment schedule is not fixed and depended on the needs of the healthcare practitioners. With the disparity of timing, the new staff may have missed the scheduled training. Nevertheless, training is crucial for new healthcare practitioners to enable them to use the system immediately to perform their work. The solution for this issue is for IT department to conduct internal training. The view is explained by a respondent who cited:

> "Previously training was limited because it was conducted by the vendor under the purchasing contract. As for pharmacist, new employment is less than doctors. We hired one to two new pharmacists at one time in which training is not worth with small amount of trainees. Instead, our staff conducted our own training. Currently, with the cooperation by IT staff, we generated our own training modules, therefore we can regularly conduct the training without having to rely on the vendor." (Respondent 4)

Based on the interviews findings, sufficient training led to competent HIS users in performing their job. Mismatch vendor schedule with organization new recruitment led to additional costing to the organization.

5.5 Task-Related Stressor

Task-related stressor refers to the task-related factors that affect the individual and require extra coping strategies [25]. This includes time pressure, high workload and interruptions. The task-related stressors may be resulted in selecting the wrong drug, giving wrong dosage, and overlook important information. Three respondents expressed this inference in the following ways:

> "Doctors who performed their task in a hurried manner may unintentionally select the wrong drug provided in the list. For an example, a doctor intended to order calcium lactate inadvertently selected calcium carbonate instead in the medication list." (Respondent 4)
> "Busy doctors overlooked some information and ended up making some errors." (Respondent 1)
> "90 % of the medical errors were due to human error. Take for example, a doctor ordered 5 mg medication, but a pharmacist dispensed 15 mg instead of 5 mg. Carelessness happened due to patient interruptions and press for time is the major reason for human error." (Respondent 3)

The interviews revealed that the presence of task-related stressor led to negative consequences to the HIS usage. Pressed for time, high workload, and interruption may increase the probability of committing errors.

5.6 HIS Safety Use Procedure

Procedure is useful for creating standard and guideline for ensuring the safety use of HIS. However, the interview indicated that here is no specific procedure for safe use of HIS. One respondent remarked that:

> "To my knowledge, there is no policy or procedure on the safe use of HIS". (Respondent 3)

On the other hand, security mechanism such as access control based on role has been applied to allow only authorized users to view or edit the content provided by HIS as reported by one interviewee.

> "Currently ISO, or guideline on the HIS safety is not available. Instead, we refer to our organization Information and Communication Technology (ICT) Security Policy which is more focus on the security aspect such as login procedure". (Respondent 6)

In the same view, one respondent added that despite the absence of HIS safety use procedure, healthcare practitioners are aware that they should not jeopardize patient safety. The respondent conveyed that:

> "As far as I know, there is no safety standard with regards to the use of HIS. Nevertheless, healthcare practitioners are aware that they should not give wrong medication, or follow wrong procedure. On the other hand, there is standard for security". (Respondent 5)

The interview data discovered that procedure for ensuring the safe use of HIS is not well defined and prepared. The current safety procedure is concerned on the security aspect of the HIS. Therefore, insufficient guideline available for ensuring the safety use of HIS. Hence, all the themes derived from this preliminary study are depicted in Table 2.

6 Discussion

This preliminary study identified key themes describing the usage of HIS. The interviews findings demonstrate that the usage of HIS can result in some technology-induced errors. Healthcare practitioners of this study attributed that the

Table 2. Summary of themes, explanation, and implication

Themes	Explanation	Implication
The use of HIS resulted in errors	The use of HIS may resulted in errors.	Interaction between HIS and user may cause error such as wrong patient, medication, dose, and x-ray result.
System quality	System usability and compatibility are the measures for system quality that influence on the safety HIS usage.	• Poor system usability such as complicated interface design, and illogical ordering of the medication terms in HIS interface inhibit healthcare practitioners to use HIS efficiently, and potentially to make errors. • Insufficient functions that are required to execute crucial clinical process vulnerable for errors to be occured. • Inappropriate system integration potentially pose risk to patient safety.
Information quality	Information relevancy is the measure for information quality that influences on the safety HIS usage.	Irrelevant alerts may cause healthcare practitioners to ignore the alerts and fail to act appropriately.
Training	Training influence healthcare practitioners to use HIS appropriately.	Inadequate training may cause healthcare practitioners to spend more time in using HIS and unable to use the HIS properly.
Task-related stressor	Task-related stressor includes time pressure, high workload, and interruptions influence on the safety HIS usage.	Time pressure, high workload, and interupptions may result in healthcare practitioners to perform task hurriedly, overlook at some information, and subsequently make errors.
HIS safety use procedure	Procedure for ensuring the safe use of HIS is not well defined and prepared	No specific procedure for safety use of HIS may lead healthcare practitioners to device their own way of using HIS that may potentially result in errors.

most common errors related to the HIS usage appeared to be: (1) patient's information was wrongly viewed, (2) medication was wrongly prescribed, (3) dosage of medicine was wrongly entered, and (4) laboratory result was wrongly matched with corresponding patient. Several studies examined the consequences of healthcare computerized systems revealed similar findings [16, 26]. The most repeatedly reported types of errors in community pharmacies include wrong medication, wrong dosage, wrong dose directions, and wrong patient [16]. The finding is also in line with a study that evaluated critical incidents in a tertiary care [26]. Their study reported that medication error such as wrong dosage, mixing up medication and patients were the most predominant errors.

In this study, the errors related to the HIS usage were the results of inappropriate interaction between HIS and human. The finding is supported by an analysis of electronic health record (EHR), reported that 74 of the safety concerns involved unsafe technology and 25 involved unsafe use of technology [15]. The finding is also consistent with a report mentioned that human-errors were the most frequent contributed to critical incident [26]. Besides, a study on computer-related patient safety incidents reported that 55 % of the problem were associated to technical issues relating to computer hardware, software, or networking infrastructure, and the remained 45 % were due to human factors [27]. Therefore, HIS alone is not adequate to improve the quality of care and patient safety. Sociotechnical aspects must be considered in order to realize the potential of HIS to reduce medical errors.

Additionally, our findings showed that system quality and information quality are associated to the HIS that influenced on the HIS usage. In details, system quality is related to the usability and compatibility. According to the respondents, poor user interface design that did not meet to the healthcare practitioners' requirement caused time consuming and prevented the ease of use of the systems. The HIS design hindered multiple screens display at one time, consequently prohibited the healthcare practitioners from accessing the desired information easily. Usability is also associated to the illogical ordering of the medication terms in HIS interface. Our respondents attributed medication error as illogical ordering of medication name that listed in alphabetical order rather than common medication taken by patients. Similarly, previous studies reported that illogical ordering of terms caused the failure of healthcare practitioners to select or scroll the information in the pop-up menu, leading to the incorrect data entry [16, 28, 29]. Besides, the finding showed that poor system compatibility denoted by limited functions and system integration problem can jeopardize patient safety. Likewise, a study on the safety of electronic prescribing found that limited functionality resulted in additional tasks which were not done with paper-based system were potentially induced errors [30]. Two studies reported that systems integration problems with other systems were the primary reason for incomplete and lost requests sent to other departments such as pathology and radiology as well as for missing results [27, 31].

In terms of information quality, this study pointed out that information relevancy is the antecedent related to the safety use of HIS. Healthcare practitioners ignored alerts that were perceived as irrelevance. In line with this finding, other studies reported that healthcare practitioners ignored series of irrelevance alerts and warnings produce by healthcare computerized systems that may cause danger to patients' safety [32–34]. A study conducted by Abramson el al. [34] reported that two-third of the healthcare practitioners believed the commercial EHR with more robust clinical decision support systems was not being able to improve medication safety was mainly due to alert fatigue. The study reported that majority of alerts were overridden.

Training is another HIS safety use antecedent discovered in our study. Insufficient training was the reason for the healthcare practitioners unable to use the HIS in a way it was intended to be used. Likewise, previous studies highlighted that sufficient training is crucial to acquire efficient and safe use of healthcare computerized systems, resulting in risk or error reduction as well decrease negative emotional towards the systems [35–37]. A previous study reported that inadequate training was the most frequently mentioned factor contributing to errors related to healthcare computerized systems [38]. In addition to the knowledge on how to use the HIS, training should also emphasize on the safety practice of using HIS.

Furthermore, task-related stressor is associated to the task that influence on the HIS usage. In this study, heavy workload, time pressure, and interruptions were revealed as the antecedents that have the potential for technology-induced errors. The errors can occurred by inadvertently selecting wrong drug, giving wrong dose, and overlook some important information. Our finding is consistent with a study examined the cause of prescribing errors in English general practice [33]. The study discovered that high workload in addition to the time pressure contributed to stress and subsequently vulnerable to make error. Heavy workload promote multitasking, workarounds, or inhibit the adherence to policy and procedures [39–41]. Adelman et al. [42] reported that repeated interruptions as well the ease of switching between patients screen in the healthcare computerized systems contributed to over three quarter (81 %) of the wrong-patient electronic orders errors.

One of the approaches to ensure the safety use of HIS is through policy and procedure. However, the finding discovered that specific procedure to guide on how to appropriately use the HIS was not clearly defined and prepared. Similarly, a comparative review of patient safety initiatives for national health IT in England, Denmark and Canada reported that most of the current standard and guideline emphasis on the safe design and implementation of software [17]. There are no guide on the safe system use [17]. Moreover, Kushniruk et al. [18] reviewed HIS safety improvement in Canada, US, and England revealed similar results as [17]. The safety approaches include developing standards related to usability and interface design, certifications, and directives from regulatory [18]. Hence, there is a need for developing HIS safety use procedure to ensure the safe use of HIS.

7 Conclusion and Future Works

This on-going research presented initial findings based on a preliminary investigation. Through six semi-structured interviews conducted, it was ascertained that the use of HIS may result in errors. System quality, information quality, task-related stressor, and training were identified as the antecedents towards safety use of HIS. From the interviews, it appears that healthcare organizations may need to look further towards developing procedure on how to safely use the HIS. Currently, it was reported that no safety procedure for Malaysia HIS adoption. The result from this in-progress work will assist towards the development of a conceptual model for safety use of HIS. Multidimensional sociotechnical aspects of HIS safety use antecedents will be considered in formulating a more comprehensive model which will contribute to the body of knowledge in the HIS safety domain.

This study has several limitations. First, the study only involved voluntary interviewees. Only doctors, pharmacists, and HIS developers were the unit of analysis in this study. Therefore, the composition did not consider all levels of HIS users. Second, the interviews were conducted at a government hospital and a teaching university hospital. The study is limited by the numbers of respondents. Hence, the results of this study may not be generalized to other settings equipped with different healthcare computerized systems. Future studies should cover more respondents to obtain more views of the HIS usage in Malaysia. Moreover, mixed methods combines qualitative and quantitative research methods will be employed in future study to provide stronger evidence for a conclusion through convergence of qualitative and quantitative findings. The qualitative method is adopted to explore in-depth understanding of the factors that influence HIS safety use in Malaysia context. Subsequently, quantitative methods will be employed to gain statistical results of the varying factors that have been identified through the qualitative methods as well to generalize the result. Thirdly, the findings of this study are merely based on the perceptions of HIS users and developers. Hospital records and evidence on safety were not analyzed in this study. Observation and document review analysis are recommended for the future research to gain in depth understanding and stronger evidence on the safety HIS usage. Regardless of these limitations, the preliminary investigation had provided valuable insights on safety use of HIS particularly in Malaysia settings.

This paper demonstrated an initial effort in exploring and identifying the antecedents towards safety use of HIS. The next phase of this research will proceed with more interviews, once the ethical approval is obtained. From there on, the development of conceptual model for the safety use of HIS will proceed. The findings of this study provides some indication of the themes that could be apparent through this preliminary investigation.

Acknowledgments. This study is funded by Research University Grant (RUG) from University of Technology, Malaysia (UTM), grant number: Q.K130000.2538.09H32, and supported by the Universiti Teknikal Malaysia Melaka (UTeM) and Ministry of Higher Education (MOHE) Malaysia.

References

1. Abdullah, Z.S.: Hospital information systems implementation framework: critical success factors for Malaysian public hospitals. Curtin University (2013)
2. IOM Health IT and Patient Safety: Building Safer Systems for Better Care (2011)
3. Runciman, W., Hibbert, P., Thomson, R., et al.: Towards an international classification for patient safety: key concepts and terms. Int. J. Qual. Health Care **21**, 18–26 (2009). doi:10.1093/intqhc/mzn057
4. Borycki, E.M., Househ, M.S., Kushniruk, A.W., et al.: Empowering patients: making health information and systems safer for patients and the public. contribution of the IMIA health informatics for patient safety working group. Yearb. Med. Inform. **7**, 56–64 (2011)
5. Mohan, J., Razali Raja Yaacob, R.: The Malaysian telehealth flagship application: a national approach to health data protection and utilisation and consumer rights. Int. J. Med. Inf. **73**, 217–227 (2004). doi:10.1016/j.ijmedinf.2003.11.023

6. Bureau NPC. Annual Report 2012. 1–142 (2012)
7. Aman, A.: Clinical information systems in private hospitals. In: 15th International Conference on Advanced Communication Technology, pp 1032–1035 (2013)
8. Taib, I.A., McIntosh, A.S., Caponecchia, C., Baysari, M.T.: A review of medical error taxonomies: a human factors perspective. Saf. Sci. **49**, 607–615 (2011). doi:10.1016/j.ssci.2010.12.014
9. Alvarez, G., Coiera, E.: Interruptive communication patterns in the intensive care unit ward round. Int. J. Med. Inf. **74**, 791–796 (2005)
10. Ash, J.S., Berg, M., Coiera, E.: Some unintended consequences of information technology in health care : the nature of patient care information system-related errors. J. Am. Med. Inf. Assoc. **11**, 104–112 (2004). doi:10.1197/jamia.M1471.Medical
11. Kohn, L.T., Corrigan, J.M., Molla, S.: To Err Is Human: Building a Safer Health System. National Academies Press, Washington, DC (2000)
12. Bates, D.W., Gawande, A.A.: Improving safety with information technology. N. Engl. J. Med. **348**, 2526–2534 (2003). doi:10.1056/NEJMsa020847
13. Magrabi, F., Ong, M.-S., Runciman, W., Coiera, E.: Using FDA reports to inform a classification for health information technology safety problems. J. Am. Med. Inf. Assoc. **19**, 45–53 (2010). doi:10.1136/amiajnl-2011-000369
14. Metzger, J., Welebob, E., Bates, D.W., et al.: Mixed results in the safety performance of computerized physician order entry. Health Aff. **29**, 655–663 (2010). doi:10.1377/hlthaff.2010.0160
15. Meeks, D.W., Smith, M.W., Taylor, L., et al.: An analysis of electronic health record-related patient safety concerns. J. Am. Med. Inf. Assoc. **21**, 1053–1059 (2014). doi:10.1136/amiajnl-2013-002578
16. Odukoya, O.K., Stone, J.A., Chui, M.A.: E-prescribing errors in community pharmacies: exploring consequences and contributing factors. Int. J. Med. Inf. **83**, 427–437 (2014). doi:10.1016/j.ijmedinf.2014.02.004
17. Magrabi, F., Aarts, J., Nohr, C., et al.: A comparative review of patient safety initiatives for national health information technology. Int. J. Med. Inform. **82**, e139–e148 (2013). doi:10.1016/j.ijmedinf.2012.11.014
18. Kushniruk, A.W., Bates, D.W., Bainbridge, M., et al.: National efforts to improve health information system safety in Canada, the United States of America and England. Int. J. Med. Inf. **82**, e149–e160 (2013). doi:10.1016/j.ijmedinf.2012.12.006
19. Hesse-Biber, S.N., Leavy, P.: The Practice of Qualitative Research. SAGE Publications Inc, Thousand Oaks (2006)
20. Creswell, J.W.: Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research, 4th edn. Pearson, Boston (2012)
21. DeLone, W.H., McLean, E.R.: The DeLone and McLean model of information systems success: a ten-year update. J. Manag. Inf. Syst. **19**, 9–30 (2003)
22. Petter, S., DeLone, W., McLean, E.: Measuring information systems success: models, dimensions, measures, and interrelationships. Eur. J. Inf. Syst. **17**, 236–263 (2008). doi:10.1057/ejis.2008.15
23. Cresswell, K., Sheikh, A.: Organizational issues in the implementation and adoption of health information technology innovations: an interpretative review. Int. J. Med. Inf. **82**, e73–e86 (2013). doi:10.1016/j.ijmedinf.2012.10.007
24. Häyrinen, K., Saranto, K., Nykänen, P.: Definition, structure, content, use and impacts of electronic health records: a review of the research literature. Int. J. Med. Inf. **77**, 291–304 (2008). doi:10.1016/j.ijmedinf.2007.09.001

25. Konradt, U., Hertel, G., Schmook, R.: Quality of management by objectives, task-related stressors, and non-task-related stressors as predictors of stress and job satisfaction among teleworkers. Eur. J. Work Organ. Psychol. **12**, 61–79 (2003). doi:10.1080/13594320344000020
26. Scharein, P., Trendelenburg, M.: Critical incidents in a tertiary care clinic for internal medicine. BMC Res. Notes **6**, 276 (2013)
27. Magrabi, F., Ong, M.S., Runciman, W., Coiera, E.: An analysis of computer-related patient safety incidents to inform the development of a classification. J. Am. Med. Inf. Assoc. **17**, 663–670 (2010). doi:10.1136/jamia.2009.002444
28. Kushniruk, A.W., Triola, M.M., Borycki, E.M., et al.: Technology induced error and usability: the relationship between usability problems and prescription errors when using a handheld application. Int. J. Med. Inf. **74**, 519–526 (2005). doi:10.1016/j.ijmedinf.2005.01.003
29. Walji, M.F., Kalenderian, E., Tran, D., et al.: Detection and characterization of usability problems in structured data entry interfaces in dentistry. Int. J. Med. Inf. **82**, 128–138 (2013). doi:10.1016/j.ijmedinf.2012.05.018
30. Westbrook, J.I., Baysari, M.T., Li, L., et al.: The safety of electronic prescribing: manifestations, mechanisms, and rates of system-related errors associated with two commercial systems in hospitals. J. Am. Med. Assoc. **20**, 1159–1167 (2013). doi:10.1136/amiajnl-2013-001745
31. Zwaanswijk, M., Verheij, R.A., Wiesman, F.J., Friele, R.D.: Benefits and problems of electronic information exchange as perceived by health care professionals: an interview study. BMC Health Serv. Res. **11**, 256 (2011). doi:10.1186/1472-6963-11-256
32. Armada, E.R., Villamañán, E., López-de-sá, E., et al.: Computerized physician order entry in the cardiac intensive care unit : Effects on prescription errors and workflow conditions. J. Crit. Care **29**, 188–193 (2014)
33. Slight, S., Howard, R., Ghaleb, M.: The causes of prescribing errors in English general practices: a qualitative study. Br. J. Gen. Pract. **63**, e713–e720 (2013). doi:10.3399/bjgp13X673739
34. Abramson, E.L., Patel, V., Malhotra, S., et al.: Physician experiences transitioning between an older versus newer electronic health record for electronic prescribing. Int. J Med. Inf. **81**, 539–548 (2012). doi:10.1016/j.ijmedinf.2012.02.010
35. Nanji, K.C., Rothschild, J.M., Salzberg, C., et al.: Errors associated with outpatient computerized prescribing systems. J. Am. Med. Inf. Assoc. **18**, 767–773 (2011). doi:10.1136/amiajnl-2011-000205
36. Hincapie, A.L., Warholak, T., Altyar, A., et al.: Electronic prescribing problems reported to the pharmacy and provider ePrescribing experience reporting (PEER) portal. Res. Soc. Adm. Pharm. **10**, 647–655 (2014). doi:10.1016/j.sapharm.2013.08.007
37. Condren, M., Studebaker, I.J., John, B.M.: Prescribing errors in a pediatric clinic. Clin. Pediatr. (Phila) **49**, 49–53 (2010). doi:10.1177/0009922809342459
38. Thompson, D., Duling, L.: Computerized physician order entry, a factor in medication errors: descriptive analysis of events in the intensive care unit safety reporting system. JCOM **12**, 407–412 (2005)
39. Hakimzada, A.F., Green, R.A., Sayan, O.R., et al.: The nature and occurrence of registration errors in the emergency department. Int. J. Med. Inf. **77**, 169–175 (2008). doi:10.1016/j.ijmedinf.2007.04.011
40. Niazkhani, Z., Pirnejad, H., van der Sijs, H., Aarts, J.: Evaluating the medication process in the context of CPOE use: the significance of working around the system. Int. J. Med. Inf. **80**, 490–506 (2011). doi:10.1016/j.ijmedinf.2011.03.009

41. Cheng, C.H., Goldstein, M.K., Geller, E., Levitt, R.E.: The effects of CPOE on ICU workflow: an observational study. In: AMIA 2003 Symposium Proceedings, pp 150–154 (2003)
42. Adelman, J.S., Kalkut, G.E., Schechter, C.B., et al.: Understanding and preventing wrong-patient electronic orders: a randomized controlled trial. J. Am. Med. Inf. Assoc. **20**, 305–310 (2013). doi:10.1136/amiajnl-2012-001055

Integration of RIS and PACS for Efficient Knowledge Management in Healthcare – A Case in Slovenia

Boštjan Šumak[1(✉)], Matej Štumpfl[2], and Maja Pušnik[1]

[1] Faculty of Electrical Engineering and Computer Science, University of Maribor, Maribor, Slovenia
{bostjan.sumak,maja.pusnik}@um.si
[2] Health Centre Velenje, Vodnikova 1, Velenje, Slovenia
matej.stumpfl@zd-velenje.si

Abstract. Contemporary information and communication technology is accelerating development of new and innovative ways for disseminating and sharing health information and knowledge between different healthcare institutions and users. Lately, several initiatives and technology innovations were introduced in the field of medical informatics. Researchers and practitioners are investing a lot of time and effort in research and development of standards for interoperable healthcare knowledge delivery and efficient healthcare KM. There are several challenges related to the healthcare KM that have to be addressed such as complexity and interoperability of healthcare information systems, impact of medical errors, patients' data security, etc. In this paper, a case of integration of the RIS/PACS system with existing information systems in health centre Velenje is presented. Additionally, implementation of the teleradiology is presented together with an analysis of its impact on the radiology practice and cost savings.

Keywords: RIS · PACS · HL7 · DICOM · Healthcare · Teleradiology

1 Introduction

When it comes to a patient, his or her life quality can depend on the quality of services he or she retrieves in healthcare. Consequently, healthcare is one of the most important and crucial domains. According to the data presented by WHO, in Slovenia 9.4 % of GDP is spent on health, or 2552 USD per person [1]. For comparison, in the UK 9.6 % of GDP is spent on health, or 3480 USD per person. Because of the growing demand for better and more effective healthcare services, researchers from different health and medicine related domains are searching for new and innovative solutions in different healthcare fields. The digital era echoes in major shifts and developments of information and communication technology (ICT) resulting in innovations, which are dramatically changing the way of life. These innovations have also made a big impact on: (1) the way healthcare services are delivered and used, and (2) the relationship between healthcare providers and consumers [2].

In last years, contemporary ICT is increasingly being recognized as a significant instrument that can be used to accelerate the progress in healthcare practices.

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 722–737, 2015.
DOI: 10.1007/978-3-319-21009-4_53

For example, contemporary mobile IT/IS solutions have been recognized as emerging and enabling technologies and applied in several countries for emergency care and/or general healthcare [3]. Governments around the world are spending a lot of resources in establishing initiatives and projects (e.g. electronic health records, telemedicine, tele-radiology, tele-dermatology, e-public health care, etc.) with focus in search for innovative ways, how to use ICT to provide cost effective and quality healthcare services.

Worldwide, several initiatives and projects were established for development of specifications of different standards and practices for improving interoperability between different healthcare information systems. Clinical Information Modeling Initiative (CIMI) [4] is an international collaboration, which main mission is to improve the interoperability of healthcare systems through shared implementable clinical information models for health records, messages and documents. Semantic Interoperability for Health Network (SemanticHealthNet) [5] is a project with objective to develop a scalable and sustainable pan-European organizational and governance process for the semantic interoperability of clinical and biomedical knowledge, which should help to ensure that EHR systems are optimized for patient care, public health and clinical research across healthcare systems and institutions. Another project with focus in interoperability in healthcare is the project Antilope [6], which is a thematic network co-founded by European Union. The project comprises key national and international organizations working together in providing guidelines and best practices for converging eHealth platforms and practices. European Union is also co-founding activities in the project EXPAND (Expanding Health Data Interoperability Services) [7], which goal is to progress towards an environment of sustainable cross border eHealth services established at the European level as well as at the national level. EXPAND is expected to maintain and further develop interoperability assets with European scope by integrating all relevant recommendations and decisions taken by different projects like epSOS, SemanticHealthNet, CALLIOPE, EHR4CR and related projects. The epSOS project [8] was founded in 2008 with aim to design, build and evaluate a service infrastructure that demonstrates cross-border interoperability between EHR systems in Europe. CALLIOPE project [9] was also launched in 2008 as a thematic network with goal to produce value for decision makers, implementers, professionals, patients and other stakeholders in projects dealing with establishment of interoperable eHealth services. Another project that is active for last five years EHR4CR (EHR for Clinical Research) [10], which is aiming at providing solutions for reusing data from EHR systems for clinical research.

Contemporary ICT drives development of new and innovative ways for disseminating and sharing knowledge between different organizations. One of the biggest challenges in healthcare remains effective knowledge management (KM), where more and more healthcare organizations are starting to realize the need for adopting the KM in order to support patients with quality healthcare services. Because medical knowledge is growing with a rapid speed, it is very difficult for healthcare professionals to keep up with. Because of this gap, researchers and practitioners are looking for innovative solutions for successfully establishing the KM in healthcare processes. In latest studies several new knowledge engineering techniques were proposed and built for decision making and process management supporting clinicians, ensuring that their decisions are consistent with knowledge and clinical procedures are carried out in a timely, efficient and safe way [11].

Healthcare presents new challenges in KM e.g. the complexity of the healthcare system, impact of medical errors, secure exchange of critical patient data, interoperability, support for processes, information management, etc., which need to be well understood. One of the goals of KM in healthcare is to lower costs through more effective sharing of knowledge. In medical practice today it is normal that patients visit different medical institutions when attending medical treatments, which can be performed in dislocated facilities, cities or even countries. As patients travel through different organizations, managing the information flow between the various components becomes more complicated [12]. Consequently, the information about patient is spread across different systems and boundaries. Therefore, the basic requirement is to provide access to all the patient's medical data needed for treatment process.

As different healthcare organizational units often have their own specialized IT applications, one of the main challenges is to integrate these solutions [13]. Challenges with interoperability are not unique to the healthcare domain and were addressed and researched in many other business sectors like banking, insurance, manufacturing sector, etc. However, because of the need to consolidate and manage healthcare information in order to maximize interaction and value of it, one of the key focus research areas in KM in healthcare remains efficient data integration from existing systems and development of secure and reliable public sharing mechanisms [12].

The growth of the internet and related technology accelerated development of new scanner technology. Contemporary scanning units are able to exchange results of imaging such as x-rays, CTs, and MRIs between physically dislocated medical institutions and radiologists. Teleradiology was born, which introduced a new way of sharing results of the imaging procedures, where the remote access to imaging services is improving patient care in many ways. For example, it is very easy to get an instantaneous second opinion for a difficult diagnosis from a specialist located anywhere in the world. The use of technology for teleradiology enables to carry out more imaging procedures at non-significant increase of costs. However, for successful knowledge transmission about the patient's diagnosis between different teleradiology nodes or actors, new technologies were developed and standardized.

This paper is organized as follows. In next section, the importance of knowledge management in healthcare is summarized. In section three some of the most important standards for medical data exchange and integration of healthcare information systems are presented. Section four reports experiences and results of an integration project, which was carried out in Slovenia. In the described case different healthcare information systems and teleradiology technology were integrated in order to improve the efficiency of healthcare services. The last section concludes the paper with future work directions.

2 Knowledge Management in Healthcare

Healthcare processes are very complex in different aspects. They can involve different stakeholders, including physicians, specialists, technicians (lab, radiology, etc.), psychologists, etc. Very often third parties such as administrators, managers, healthcare ministry, insurance companies, education organizations, academics and other can be part in such processes. Partners participating in an individual healthcare process can be positioned at different geographical locations. The amount of knowledge being created by healthcare partners can be very extensive. Innovations of ICT in healthcare systems require coordination of different components: manpower, medical materials and relevant technical components. Therefore, one of the biggest challenges for healthcare organizations is building effective relations between separate islands of systems and medical units in order to enhance capabilities and maintain competitiveness [2]. This is also one of many reasons why healthcare organizations have started to establish and implement initiatives for identifying, sharing and managing different knowledge assets [14].

Health management is a knowledge intensive activity and most organizations have specialist sub-domains, each with its own vocabulary, knowledge base and software applications [12]. It is very important that any knowledge asset created by individual healthcare partner is shared to all involved partners in an efficient way. Therefore, the healthcare domain is calling for appropriate KM technologies and practices for providing quality services. KM has already been recognized in other domains as an approach that can help organizations achieve goals more quickly, more effectively and more systematically. KM in healthcare aims to improve the quality, efficiency, and effectiveness of healthcare using knowledge-based tools, techniques, and programs for system improvement. Additionaly, KM in healthcare promises cost savings, new revenue generation, innovation and return of investment (ROI). Abidi defined KM in healthcare as following [15]:

"The confluence of formal methodologies and techniques to facilitate the creation, identification, acquisition, development, preservation, dissemination and finally the utilization of the various facets of a healthcare enterprise's knowledge assets".

If applied properly, healthcare can benefit from KM through (1) improved operating performance as a result of effective knowledge sharing and best practices, (2) better decision-making and innovation adoption, and (3) enhanced competitive positioning in global healthcare business. Recent developments in ICT have had a big impact on KM in healthcare, influenced by following three elements [16]:

1. *Information Technology* – Information technology will allow every laboratory report to carry knowledge. About three million laboratory reports are produced per million population every year. Latest trends show an increased amount of digitized information data that is available 24 h a day, seven days a week [17]. Digitized information can be delivered not only to clinicians but also directly to patients [16].
2. *Knowledge* - Knowledge and the Internet interact, because the Internet allows general access to the state-of-the-art knowledge. Healthcare services have to ensure that knowledge is provided efficiently. Globalization of business has impact on healthcare delivery since it can occur anywhere in the world and the knowledge represents the true source of competitive advantage [17]. Knowledge management will be a central responsibility of health service management in the 21st century [16].
3. *Patients* – Attitudes of patients are changing and their expectations are greater than ever for: better health; better healthcare; more open decision-making; etc. [16].

Teleradiology is a technology, which is contributing to the healthcare KM in a very important way. It is removing physical boundaries and enabling to use radiology specialists from anywhere in the world in complex image reading and diagnosing processes. In this way, teleradiology is accelerating the exchange of knowledge and know-how among radiologists and other stakeholders offering high quality radiological medicine. In spite of the saving programs admitted by governments in most countries around the globe because of the recession, the teleradiology enabled radiologists to contain their actual and perceived value to patients and other stakeholders [18]. Because of these savings in the public sector, medical institutions had to sustain or even lower medical costs. The introduction of teleradiology technology didn't reflect just in improved healthcare knowledge management, but it is also significantly lowering costs in healthcare treatment processes.

3 Standards for Integration of Information Systems in Healthcare and Knowledge Delivery

One of the biggest challenges in medical informatics today is still efficient and secure exchange of information and communication. A growing tendency is in the need for providing interoperability at different levels of healthcare information systems. The reason for the interoperability need is in fact that typical architectures of healthcare institutions systems are composed of different separate architectures that were built as isolated information systems, which were developed and optimized for individual department's needs (e.g. radiology, cardiology, etc.). At the time these systems were built, there was usually no vision or need for ability to connect with other internal or external healthcare related processes. Because of the nature of the data being exchanged, information security is also one the most critical fields in healthcare informatics domain. Analysis in these fields show that in most cases healthcare informatics, at least in Slovenia, lacks significantly behind other domains such as the economy, banking, etc.

Various organizations realized the need for consolidation of the data produced by different healthcare related information systems to enable global patient-centered view and to support interoperability of intra- and inter-connected healthcare processes. These needs motivated different initiatives that started developing standards for medical data and message exchange. Standards are a prerequisite for interoperability between cross-departmental and cross-organizational healthcare processes.

In medical informatics, several technologies for medical data exchange and integration were proposed, developed and accepted as standards. Some of the standards have already been adopted in healthcare information systems, and some are still in the adoption processes. In general, medical standards can be divided into (1) commercial or payable, which are accessible for example by payment of membership in organizations developing the standards, and (2) non-commercial and/or open source standards - accessible to public, for which more and more organizations, researchers and developers are being interested. In practice, it is also very common that solutions are built on both, commercial and non-commercial standards. Among non-commercial standards, the most known and developed standard is OpenEHR. When we look for commercial standards, the most known and developed one is the HL7 (Health Level Seven) standard. Today, HL7 is the leading standard for systems integration in healthcare [13]. Following subsections provide introduction of most known standards.

OpenEHR. OpenEHR virtual community provides open source standards on interoperability and compatibility in e-health. OpenEHR is mainly focused in development of specifications for EHRs and systems. These specifications include [19]: (1) a health information reference model, (2) a language for building clinical models or archetypes, based on extensible markup language (XML) and archetype definition language (ADL), and (3) a query language - archetype query language (AQL). Solutions that conform openEHR standards are open in terms of data (they are valid to the published openEHR XML Schemas), models (written based on ADL) and APIs. Most important contribution of openEHR efforts is in providing standards for systems and tools for computing with health information at a semantic level and enabling medical analytics functionality for decision support and research activities. An abstract architectural model of such solutions is presented in Fig. 1.

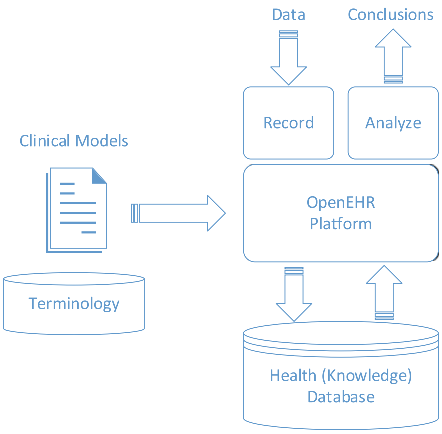


Fig. 1. OpenEHR abstract architecture [19]

Efforts of the openEHR community resulted in development of international, online clinical knowledge resource, called Clinical Knowledge Manager (CKM). CKM is a system for collaborative development, management and publishing of a wide range of clinical knowledge resources (archetypes in XML or ADL format, templates, artefact release sets, metadata relating to clinical models, etc.). Archetypes can be used for example for modeling of actions (clinical intervention, imaging examination, etc.), evaluation (clinical doc, clinical report, clinical synopsis, etc.), observation (weight, blood pressure, body temperature, eye examination, etc.), etc.

In general, CKM provides services for: (1) a library of clinical knowledge artefacts - currently openEHR archetypes and templates; (2) supporting the full life cycle management of openEHR archetypes through a review and publication process; and (3) governance of the knowledge artefacts. Basically, CKM enables the implementation of knowledge governance within and across the health enterprise. Along with the internationally open and available CKM - openEHR CKM (http://www.openehr.org/ckm/), several instances of CKM were established all over the world as a result of local initiatives, like NEHTA CKM,[1] UK Clinical Models,[2] Norwegian CKM,[3] and Slovenian CKM.[4] In the Slovenian CKM instance, 481 archetypes have been published so far (see Fig. 2).

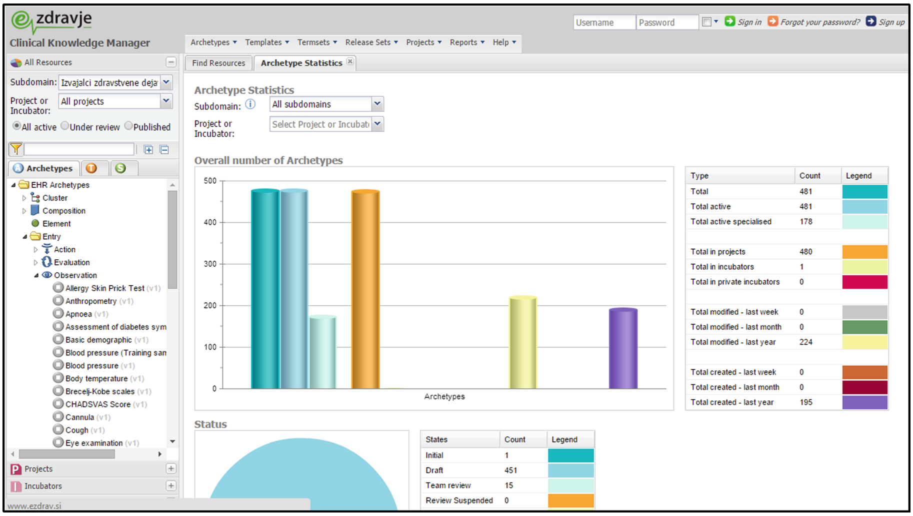


Fig. 2. Slovenian Clinical Knowledge Manager

CEN/ISO EN13606. The EN13606 is a European norm from the European Committee for Standardization (CEN) and also approved as an international ISO standard. Main purpose of the standard is to enable semantic interoperability in the electronic health record communication [20]. Main goal of the standard is to define information architecture for communicating electronic health record (EHR) data of a single patient between different systems included in the healthcare process, like different EHR systems, health knowledge management repositories, clinical applications, decision support components, etc. The standard makes a clear separation between information and knowledge through the dual model architecture: reference model and archetypes. Information is structured using a reference model that contains the basic entities for representing any information of the EHR. The EN13606 reference model is an object-oriented model that contains (1) a set of primitive types, (2) a set of classes for defining the building blocks of EHR, (3) a set of auxiliary classes for describing the context information that needs to be attached to an EHR, and (4) classes for describing

[1] http://dcm.nehta.org.au/ckm/ - hosted by NEHTA for the Australian national eHealth program.

[2] http://www.clinicalmodels.org.uk/ckm/ - the CKM for resources developed collaboratively within UK projects.

[3] http://arketyper.no/ckm - hosted and managed by the Norwegian national eHealth program.

[4] http://ukz.ezdrav.si/ckm/OKM.html - hosted and managed by the Ministry of Health in Slovenia.

demographic data and communication with EHR fragments. The health knowledge is constructed on archetypes or formal definitions of clinical concepts (e.g. discharge report, glucose measurement or family history) in the form of structured and constrained combinations of the entities defined in the reference model.

Health Level Seven, HL7. Health Level Seven, Inc (HL7) [21] is a non-profit international organization for development of standards dedicated to providing framework for exchange, integration, sharing, and retrieval of electronic health information that supports clinical practice, delivery and evaluation of health services. HL7 provides standards for interoperability that improve care delivery, optimize workflow, reduce ambiguity and enhance knowledge transfer among different stakeholders (e.g. healthcare providers, government agencies, patients, etc.).

HL7 v2. Up to version 2.6, the HL7 standard for EHR mainly focuses on exchange of data between medical information systems [22]. HL7 standards versions 2.x were developed for formalizing electronic processes like administrative, logistical, financial and clinical processes in healthcare. Despite the formalization there were still problems with interoperability between different healthcare systems, because of the lack of explicit information model in v2.x messages [22]. For efficient message exchange between different healthcare systems detailed agreements had to be provided. HL7 v2.x messages are ASCII messages with records arranged in rows. Records are divided into segments or fields.

HL7 v3. To overcome problems with interoperability, HL7 standard v3 was introduced in 1995 and finally published in 2005. The new standard was based on (1) an object-oriented data model called Reference Information Model (RIM) and (2) XML based Clinical Document Architecture (CDA). The RIM is used to express the data content and to represent semantic and lexical connections between different pieces of information. The CDA, which lies on top of RIP provides encoding, structure and semantics of clinical documents [22].

FHIR. HL7 created a new framework called Fast Healthcare Interoperability Resources (FHIR). FHIR combines the best features of HL7 v2.x, v3.x and CDA product lines while leveraging the latest web standards and applying a tight focus on implementability [23]. FHIR can be used in different context like mobile phone applications, cloud communications, EHR-based data sharing, etc. The FHIR framework also provides out-of-the-box solutions for development of interoperable solutions based on web standards like XML, JSON, HTTP, Atom, OAuth, etc. Therefore, FHIR is a major step toward enabling development of RESTful architectures supporting seamless exchange of healthcare information using messages or documents.

4 Integration of RIS and PACS in a Healthcare Organization – a Case in Slovenia

With the advent of new technologies and healthcare delivery practices, the need for usable and easily accessible and integrated information related to healthcare data is bigger than ever [12]. In the healthcare industry, different health related organizations developed information systems that were used in an island mode. Integration projects that aim integrating legacy healthcare systems are usually confronting with challenges related to interoperability and enormous costs related to maintenance of the integration links. Development of process-oriented information systems for supporting healthcare practices can trigger a lot challenges. Such solutions have to provide capabilities for efficient healthcare process support, information and knowledge management at different levels. Information systems that were built in the past hardly meet such requirements, which has led to new ways for integration of heterogeneous applications [13].

Lately, medical institutions have started to employ new Radiology Information Systems (RIS) and Picture Archiving and Communication Systems (PACS), mostly because existing analog X-ray devices are obsolete and have to be replaced with digital technology. RIS is a computer system that stores and processes the information for a radiology department and can be linked to the hospital or medical information system [24]. RIS is being used for planning, monitoring and communication of all data regarding patients and its investigations in the radiology. The RIS must provide correct images at the correct time to the correct users. PACS is a system that can be used to communicate and archive medical imaging data, mostly images and associated textural data generated in a radiology department and disseminated throughout the hospital [25]. A PACS is usually based on the Digital Imaging and Communications in Medicine (DICOM) standard. DICOM is an industry standard for transferring radiologic images and other medical information between computers [26]. DICOM is a standard that enables interoperable digital communication between diagnostic and therapeutic equipment and systems built by different manufacturers.

The most sophisticated products provide an out-of-the box solution that is already uniting both - RIS and PACS capabilities in one product. However, there are also separate products for RIS and PACS available, which if employed, usually require an extra integration project.

The health centre Velenje in Slovenia has employed a healthcare information system (HIS), called IRIS. HIS systems in Slovenia are rapidly evolving and becoming very complex. One of the reasons are fast changes in requirements made by the health insurance institute of Slovenia (ZZZS). The IRIS system is responsible for supporting healthcare processes and practices like management of patients and their visits, management of patient's diagnosing information, etc.

The health centre Velenje has also employed a RIS/PACS system called ISSA. The ISSA system was installed on a virtual server in a virtual cluster. At the time of employment of the ISSA system, four measurement instruments were linked with the new RIS/PACS system (see Fig. 3): a new x-ray machine, an older analog x-ray machine with the support of CR, a digital dental x-ray machine, and an ultrasound machine. All connected instruments support the DICOM standard, which was used for integration with the ISSA system. Although DICOM is a standard technology, there were some challenges in the integration process. The problem was due the data provided by instruments, which has some extra attributes and values specific to the manufacturer and not to the DICOM standard.

Employment of the ISSA system required effective integration with the IRIS system. These two systems have to be integrated in order to provide efficient support for automated processing of healthcare practice and to deliver right patients' data to right

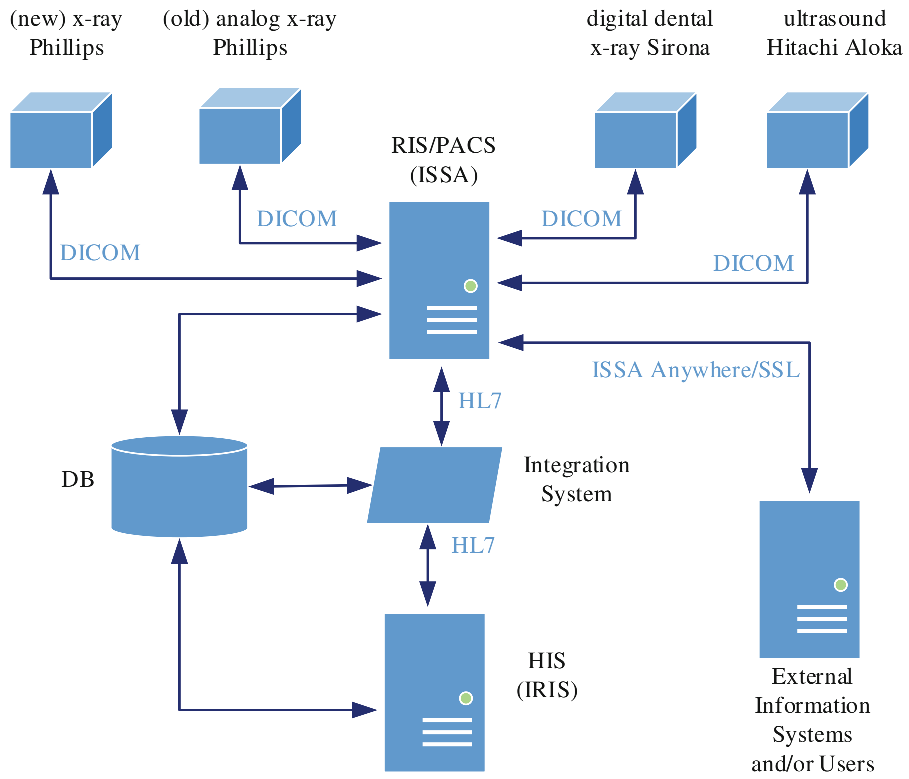


Fig. 3. A schematic view of the integration of different devices and systems with the RIS/PACS

people in an efficient way at minimal costs. For example, the IRIS system has to be able to order imaging a patient in the radiology department. Basically, the IRIS system forms an electronic order (e-order) containing data about the patient and about what has to be imaged. When completed, the e-order is sent to the radiology department. At the radiology department, a radiology engineer can find the e-order in the device's worksheet. Until the engineer proceeds with the imaging, the e-order stays open and can be modified (e.g. new types of imaging can be added, existing modified and/or deleted, order can be canceled, etc.). The e-order can be modified in both systems - in the RIS/PACS and in the HIS system.

The communication between the ISSA and IRIS was implemented at two levels (see Fig. 3): (1) a database, providing tables through which both systems exchange data, and (2) integration system, that is responsible for creating and sending HL7 messages based on the data in the exchange tables. When an e-order is triggered in the IRIS system, stored procedures are executed in the database system that are responsible to record information about the order (e.g. name of the patient, identification number of the treatment, date of birth, type of imaging, data about the doctor, etc.). Next, the integration system reads the data and creates an HL7 message that is sent to the ISSA system. An example of such HL7 v2 message is shown in Fig. 4 and an XML version in Fig. 5. The ISSA system has to process the received HL7 message and constructs a DICOM message that is forwarded to the correct instrument for imaging. The integration system is also responsible for updating status of the e-order in the exchange tables, so the status of the order can be checked by the IRIS system as well.


```
MSH|^~\&|IRIS||INFOSISTEM||20150209033048||ORM^O01|000000
0014591409|P|2.3|
PID|||2000000000000051396|11111111|ŠTUMPFL^MATEJ||1972022
2|M|||SPODNJI RAZBOR 61 B^^Velenje^^3320^Slovenija||||
PV1||||||||546^^|555
ORC|NW|0000000014591409|0000000014591409||SC||||20150209|
||11134^NOVAK|||||_02_02^_02_02-SERVISI_ZOBNI
RTG|01^Nedodeljena naprava
OBR|1|0000000014591409|0000000014591409|9000^Ortopan||201
50209|20150209033048|||||||||0^Unassigned|||||||||S||1|||
|^^|||||20150209033048|
```


Fig. 4. Example of an HL7 v2 message for dental imaging (ortopan)


```xml
<?xml version="1.0" encoding="UTF-8"?>
<result>
  <exchangehead_status>0</exchangehead_status>
  <exchangehead_adress>SP.RAZBOR 61</exchangehead_adress>
  <exchangehead_id>15316</exchangehead_id>
  <exchangehead_refferingcode>11134</exchangehead_refferingcode>
  <exchangehead_department>_02_02</exchangehead_department>
  <exchangehead_an>0000000014591409</exchangehead_an>
  <exchangehead_firstname>MATEJ</exchangehead_firstname>
  <exchangehead_lastname>ŠTUMPFL</exchangehead_lastname>
  <exchangehead_comment/>
  <exchangehead_refferingdesc>NOVAK</exchangehead_refferingdesc>
  <exchangehead_city>3320^Velenje</exchangehead_city>
  <exchangehead_country>Slovenija</exchangehead_country>
  <exchangehead_patientcode>11111111</exchangehead_patientcode>
  <exchangehead_refferalcode/>
  <exchangehead_sex>M</exchangehead_sex>
  <exchangehead_departmentdesc>02_02-SERVISI_ZOBNI RTG
  </exchangehead_departmentdesc>
  <exchangehead_patientaltcode>2000000000000051396
  </exchangehead_patientaltcode>
  <exchangehead_datebirth>1972-02-22 00:00:00.0
  </exchangehead_datebirth>
</result>
```


Fig. 5. Example of an HL7 v2 message transformed into XML format

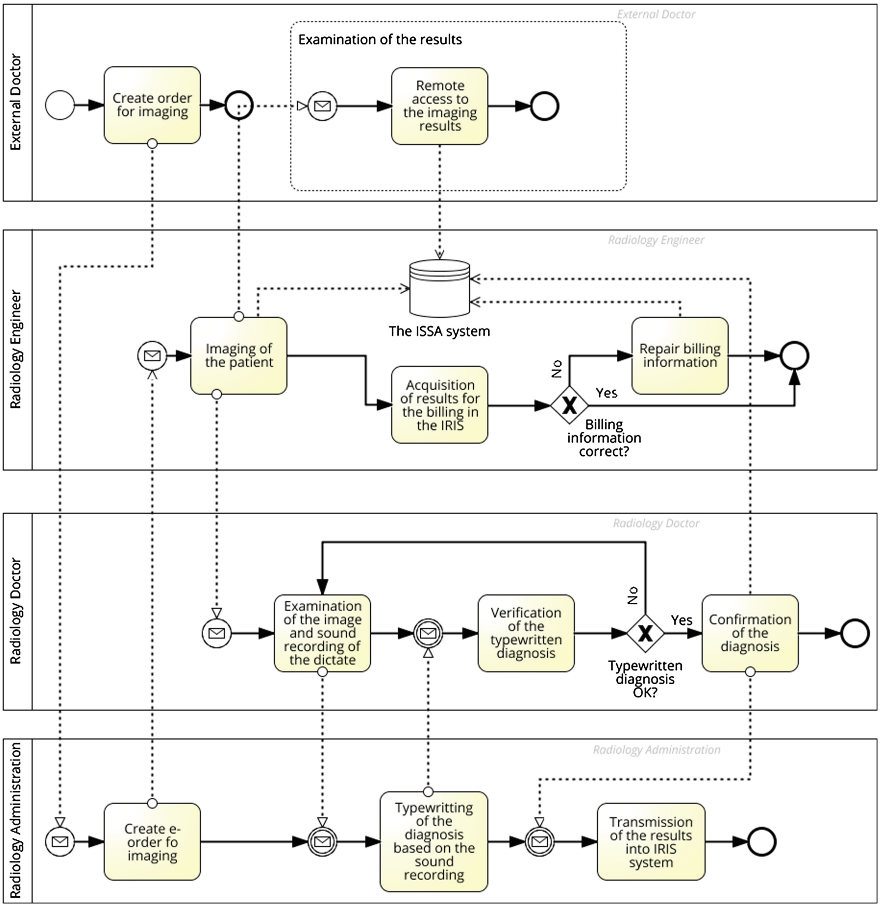


Fig. 6. The process of processing of the scanning results in the radiology department

The processing of a patient involves following steps (see Fig. 6):

1. Imaging of the patient and allocation of the image to the radiology doctor for examination
2. Examination of the image and sound recording of the dictate
3. Typewriting of the diagnosis based on the sound dictate
4. Verification of the typewritten diagnosis
5. Confirmation of the diagnosis
6. Transmission of the results into the IRIS system.
7. Completion of the imaging process.

After the processing in the ISSA system is completed, the IRIS system provides access to the results of the processing. To ensure a greater control, the acquisition of the results can be done only manually.

4.1 Teleradiology Implementation

In general, the teleradiology in practice means transferring radiological images and results to a remote location, which can result in better diagnosis and consequently higher quality patient care and result of the treatment. It is a technology that enables access to images and linked materials remotely. The teleradiology is opening possibilities for new practices and remote consultancy in the diagnosis processes. A radiology doctor located in the health centre has a direct access to the imaging results in the IRIS system.

However, one of the requirements when implementing the RIS/PACS system was to enable external users (e.g. doctors, dentists, and related personnel) a secure and efficient remote access to the ISSA system. It was important to enable remote reading of the images and to add a diagnosis. External users must have an instant remote access to the imaging results for punctual decisions about further treatment of their patients that were examined in the selected radiology department. It is crucial for a doctor to have an instant access to the imaging results without the need to wait for a dictate or even a physical print of the image. Before implementing the teleradiology, a dentist had to wait for results that were sent to the dentist by email, CD or printed on a foil.

To achieve remote access to the ISSA system to authorized users, a module called IssaAnywhere was developed. The IssaAnywhere module provides an instant remote access to the results of the imaging process in the ISSA system. Additionally, through this module external users are able to add diagnosis and other observations to results of examination stored in the ISSA system. Because of the nature of the information being exposed over the public network, it was essential to ensure a sufficient level of security. To provide communication security between the ISSA system and the client, the communication was implemented using the Secure Socket Layer (SSL) protocol. According to the law regulation, every patient has legitimate rights to get radiology images provided be on any media. In the health centre Velenje, the teleradiolgy is being used every day by the radiology doctor that in most cases is remotely reading images. Every year, in the health centre Velenje more than 25.000 scans are carried out. For recording a dictate while reading the image, the radiology doctor is using a special custom programmed dictaphone.

Before the digitalization of the radiology process, every image was printed out using a special printer on a foil. With introduction of the new ISSA system, images are burned on a CD. Regardless of how many scans are made for one patient in the imaging process, only one CD is recorded containing all images that were made in the radiology department. According to the data we got from the IRIS system, we conducted a cost reduction analysis for a period of one year (between 1.5.2013 and 30.4.2014). The cost reduction analysis is presented in Table 1. Approximately 1071 foil printings in a month were made. The cost of one printing on a foil is 2.20€, causing approximately 2358€ of costs per month just for printing of the images. The cost of one CD is 0.27€. Accordingly, the average cost per month for recording image on a CD is 289€. In summary, the health centre Velenje is saving approximately 2068€ per month by replacing classical printing on a foil with a CD recording of the image. Together in a year period, these cost savings were 24825.59€.

Table 1. Costs reduction analysis related to RTG image digitalization

	Number of foil printings	Printing on a foil cost	Cost of CD recording	Saving
May 13	1123	2470.6	303.21	2167.39
June 13	828	1821.6	223.56	1598.04
July 13	987	2171.4	266.49	1904.91
August 13	797	1753.4	215.19	1538.21
September 13	1205	2651	325.35	2325.65
October 13	1229	2703.8	331.83	2371.97
November 13	1179	2593.8	318.33	2275.47
December 13	778	1711.6	210.06	1501.54
January 14	1538	3383.6	415.26	2968.34
February 14	997	2193.4	269.19	1924.21
March 14	883	1942.6	238.41	1704.19
April 14	1319	2901.8	356.13	2545.67
TOGETHER	12863	28298.6	3473.01	24825.59

5 Conslusions and Future Work

It is no doubt that contemporary ICT has a very important role in accelerating the development of new and innovative ways for disseminating and sharing healthcare information and knowledge between different healthcare institutions. In this paper we presented several initiatives and technology innovations that were lately introduced in the field of medical informatics. Researchers and practitioners in healthcare are investing a lot of time and effort in development of standards for interoperable healthcare knowledge delivery and efficient healthcare KM. Related to the nature of the healthcare there are several challenges that have to be addressed like the complexity of healthcare information systems, impact of medical errors, patient data security, etc. One of the goals when implementing the KM in healthcare is also to provide access to patient's data anywhere and anytime at lower costs.

In this paper, some of the most known and developed standards in the field of medical informatics like EN13606, OpenEHR, and HL7 were presented. However, main focus of the paper was introduction of the teleradiology technology and its benefits in the healthcare practices in the health centre Velenje. The health centre Velenje has invested a lot of effort in implementation of a new RIS/PACS system and integration of the RIS/PACS with existing in-house information systems. The integration was implemented using standards like HL7 and DICOM. To enable teleradiology, a special module called IssaAnywhere was implemented. Through this module external doctors are able to remotely access results of the radiology scans that are being made in the health centre Velenje in a secure way.

Every year, more than 25000 scans are made using instruments in the health centre Velenje. After the digitalization, the health centre Velenje is saving approximately 2068€ per month just by replacing classical printing of the image on a foil with recording of the image on a CD. According to the data we retrieved from existing information system, in one year, the health centre Velenje has saved 24825.59€. But, saving money is just one benefit of the introduction of new RIS/PACS and teleradiology technology. There are also other benefits that we plan to investigate in our future work, like improved efficiency in imaging and diagnosing processes (e.g. ability for faster diagnosis with a second opinion), improved management of the images and corresponding information (e.g. dictates), improved patient experience and satisfaction, etc. In our future work, we plan to conduct an extensive and detailed impact analysis of these solutions in the health centre Velenje. We have started with a research, which aims to analyze how new information systems and technology affect end users (e.g. radiology doctors, administrative staff, external doctors, etc.) and their patients.

References

1. Global Health Observatory Data Repository. http://apps.who.int/gho/data/node.main
2. Wu, J.H., Chen, Y.C., Greenes, R.A.: Healthcare technology management competency and its impacts on IT-healthcare partnerships development. Int. J. Med. Inform. **78**, 71–82 (2009)
3. Wu, J.H., Wang, S.C., Lin, L.M.: Mobile computing acceptance factors in the healthcare industry: A structural equation model. Int. J. Med. Inform. **76**, 66–77 (2007)
4. Clinical Information Modeling Initiative. http://www.opencimi.org/
5. SemanticHealthNet. http://www.semantichealthnet.eu/
6. Antilope Project - Advancing eHealth InteroperabilityAntilope Project. http://www.antilope-project.eu/
7. Expand - Expanding Health Data Interoperability Services. http://www.expandproject.eu/
8. epSOS: epSOS - European Patients - Smart open Services. http://www.epsos.eu
9. Calliope - a Euorpean thematic Network for eHealth Interoperability. http://www.calliope-network.eu/
10. EHR4CR - Electronic Health Records for Clinical Research. http://www.ehr4cr.eu/
11. Fox, J., Gutenstein, M., Khan, O., South, M., Thomson, R.: OpenClinical.net: a platform for creating and sharing knowledge and promoting best practice in healthcare. Comput. Ind. **66**, 63–72 (2015)
12. Orgun, B., Vu, J.: HL7 ontology and mobile agents for interoperability in heterogeneous medical information systems. Comput. Biol. Med. **36**, 817–836 (2006)
13. Lenz, R., Reichert, M.: IT support for healthcare processes - premises, challenges, perspectives. Data Knowl. Eng. **61**, 39–58 (2007)
14. Yun, E.K.: Predictors of attitude and intention to use knowledge management system among Korean nurses. Nurse Educ. Today **33**, 1477–1481 (2013)
15. Abidi, S.S.R.: Knowledge management in healthcare: Towards "knowledge-driven" decision-support services. Int. J. Med. Inform. **63**, 5–18 (2001)
16. Gray, M.: Making the future of healthcare (2008)
17. Perrott, B.E.: A strategic risk approach to knowledge management. Bus. Horiz. **50**, 523–533 (2007)
18. Li, K.C., Marcovici, P., Phelps, A., Potter, C., Tillack, A., Tomich, J., Tridandapani, S.: Digitization of Medicine: How Radiology Can Take Advantage of the Digital Revolution (2013)
19. openEHR. http://www.openehr.org/home

20. The CEN/ISO EN13606 standard. http://www.en13606.org/the-ceniso-en13606-standard
21. Health Level Seven International, HL7. http://www.hl7.org/
22. López-Nores, M., Blanco-Fernández, Y., Pazos-Arias, J.J., García-Duque, J.: The iCabiNET system: Harnessing electronic health record standards from domestic and mobile devices to support better medication adherence. Comput. Stand. Interfaces. **34**, 109–116 (2012)
23. FHIR. http://www.hl7.org/implement/standards/FHIR-Develop/index.html
24. Radiology Information System – RIS. http://www.mr-tip.com/serv1.php?type=db1&dbs=Radiology+Information+System
25. Picture Archiving and Communication System – PACS. http://www.mr-tip.com/serv1.php?type=db1&dbs=Picture Archiving and Communication System
26. Digital Imaging and Communications in Medicine – DICOM. http://www.mr-tip.com/serv1.php?type=db1&dbs=Digital+Imaging+and+Communications+in+Medicine

Knowledge Management in Education and Research

Knowledge Management and Intellectual Capital in a University Context

Luis Leonardo Rodríguez Bernal[1], Víctor Hugo Medina García[1](✉), and José Fernando López Quintero[2]

[1] District University "Francisco José de Caldas", Bogotá, Colombia
{llrodriguez, vmedina}@udistrital.edu.co
[2] ECCI University, Bogotá, Colombia
lopez.jose@ecci.edu.co

Abstract. The knowledge management and its intellectual capital is fundamental in the university organizations where little has been implemented. The objective of this article is to keep on contributing with a proposal of measurement of the intellectual capital, from the knowledge concept, the knowledge management to its relationship with the intellectual capital in a university context by a bibliographical review. In the same way, it is explained how the intellectual capital is constituted by different types of capital which are interrelated. At the same way, the intellectual capital is presented with its elements and variables. Later, the mathematical equations are formulated for different types of capital, at the same way, a simulation is applied, an analysis is done by means of some of the main components. Finally, it is suggested the improvements of the intellectual capital, in order to optimize the knowledge management, in a university atmosphere.

Keywords: Knowledge · Intellectual capital · Knowledge management · University · Measurement

1 Introduction

Knowledge is understood as the set of experiences, knowledge, values, information, perceptions and ideas, created in the mental structure of an individual who evaluates and incorporates those elements based on their experience in order to achieve the main objective of the development of a country [1].

Taking into account the previously mentioned ideas, the success of the initiatives of the knowledge management at the organizational level (University) depends on people (teachers) with initiatives that use knowledge, in order to get results that add a value into the labor (research). This could be gotten within a model that is based on the three dimensions such as: conceptual, contextual and operational. The focus within this model is - why - who, know - where, know - when, and t – how the knowledge is improving, understanding and trying to have a more concrete vision with a greater business application. Innovation and efficiency provide an approach of benefits and productivity within the model. It needs to be moved from the individual to the collective [2].

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 741–753, 2015.
DOI: 10.1007/978-3-319-21009-4_54

Finally, the intellectual capital (IC) is defined as "the set of abilities, skills and knowledge of the people who generate value either for the scientific communities for universities, for organizations and for the society in general". In this regard, Bueno refers to the conceptual perception of Sveiby and Stewart who expressed that the intellectual capital is turned into patrimony for the companies and nations, which is reflected into two dimensions, which are: static dimension (value of the intangible in a given period of time) and the dynamic dimension (actions based on application of the intellect, such as increasing and controlling) [3].

2 Theoretical Foundation

The knowledge management (KM) considers the productive and competitive aspects named as intellectual capital (IC), which can be counted by constituting as a competitive advantage for the organizations, whose managers have taken the knowledge models using methods that involve this capital. As a result, organizations (business and academic) are concerned about retaining immersed (intellectual capital) talent in their companies [4].

This KM presents a difficulty by the constant process of changing, generating a depreciation of knowledge in economic and accounting terms [5]. Therefore, the managers of the organizations are concerned about these cycles of changes that generate a high impact in all processes that you want to carry out, within the development of organizations [6].

In addition, KM organizations have evolved, in terms of a series of features conceptions that are constituted in the following way: first conception: it has a focus related with to the system engineering, where the KM is a cycle of an administration and information processing for being used within any organization, in this way, through these mechanisms of assimilation and recruitment it is possible to present practical solutions that seek to generate a new knowledge much more solid and concrete. In addition, it is established a series of technological innovations, which are closely related to the KM, where the focus is on the information structured, rather than in the design of the knowledge of persons [7, 8]. Second conception as a process: the process of managing knowledge for satisfying the needs by identifying and exploiting resources, in order to achieve the organizational goals [9].

In addition, it is considered the systematic process of the Organization in order to achieve the success through the creation, recruitment and the behavior of knowledge [10]; At the same time, the process of managing explicitly non-material assets, which exists with the main purpose of making that the company can generate, search, store and transfer the knowledge, in this way, it can be gotten to increase productivity and competitiveness [11]. The KM includes a logical cycle that begins with the identification, creation, capture, sharing, storage and transfer of knowledge (tacit or explicit) by generating competitiveness and efficiency at the level of organizations, with the main aim of achieving corporate objectives. Third concept: the human resource through the intellectual capital (IC), where there is a differentiation of the management processes of knowledge flows related with the systems of values, which take place in the organizations. Thus, the KM is considered as a set of disciplines of management that presents the IC as an asset of the company by using technological tools and components that allow to share the knowledge in order to achieve the specific objectives of the Organization, which recognizes and shows the importance of joining efforts to achieve the primary goals in the knowledge.

In addition, the knowledge management in organizations has developed a number of generations which keep a close relationship with the previously addressed conceptions that are described in the following way: the first generation: where the knowledge is captured in other words, the information management is in favor of carrying out actions in the knowledge (to store it, to sort it and to spread it), then, It is presented in the form of standardization by means of databases, languages of interoperability, among others, for the exchanging data and electronic information in particular. The second generation: This is constituted as a method in order to identify, to codify, to structure, to store, to reuse and to disseminate the community experiences that will enrich and enable, a much more effective development of the organizational purposes. The third generation deals with the capitalization of knowledge with an organizational base which contains the communities of practice for monitoring, learning and innovation, by transforming it into a system of exchanging internal-external knowledge, in other words it is converted into a network of collaborations. Finally, it is possible to find the fourth generation: is a process of dynamic relationships between external and internal knowledge of customers, in order to generate a value. Consequently, organizations with cybernetic model (self-organization and self-governance) are based on knowledge and try to answer to the possible temporary associations [12].

3 Knowledge Management and Intellectual Capital

The KM is based on the value of intellectual capital (IC) individuals or groups decision-makers that have data, information and knowledge which allows to achieve a much more solid level of effectiveness and efficiency in organizations, through a series of innovations, thus maximizing performance and minimizing costs, which means a much more effective and sustainable, development level.

Then, the main organizational models of KM, present a specific classification regarding the IC, in the following way (Table 1.):

Following the presentation of some of the hybrid models of knowledge management with its respective components and definitions: the model E.O.SECI: adapts components of intellectual capital: human capital, structural capital and relational capital. Its contributions to the measurement of intellectual capital are specific indicators from an organizational perspective and his strategic assessment with its complex methods. Meridium Model is mainly focused on the measurement of intangible variables, to understand and improve innovation management. The EKMF model: its objective is to explore the European consensus on terminology, methodology and implementation of the management of the knowledge, in order to build a unique and balanced academic actor with exemplary network and business [13].

Therefore, it can be concluded that these models are relied on intellectual capital (IC) which is made up of other types of capital such as: human capital (Hc), relational capital (Rc) and structural capital (Stc) [14].

Table 1. Organizational knowledge management models. source: [13].

Organizational knowledge management models		
Knowledge management models	**Models of Intellectual Capital measurement**	**Hybrid models of knowledge management**
Knowledge creation (Nonaka y Takeuchi, 1995)	*Balanced Bussiness Scorecard*(Kaplan y Norton,	***E.O.SECI Model*** **(Bueno, 2001)**
Model of KPMG Consulting (Tejedor y Aguirre, 1998)	*Scandia Navigator* (Edvinsson y Malone, 1996)	**Meritium Model (Unión Europea, 2001)**
Arthur Andersen (A. Andersen y APQC, 1999)	*Intangible assets monitor* (Sveivy, 1997)	***Global and Integral Model*** **(Peña, 2002)**
KMAT Model (Andersen, 1999)	*Intelect Model* (Euroforum, 1998)	***EKMF Model*** **(Unión Europea, 2003)**
	Intelectual Capital (Drogonetti y Ross, 1998)	***e-GIC Model*** **(Lombardo, 2004)**
	Intangible Capital (Bueno, 1998)	***GC-U Model*** **(Medina, 2014)**
	Nova Model (Comunidad Valenciana, 1999)	***Other Models (Thesis)*** **(2004 - 2015)**

Therefore, human capital (Hc) is framed in the knowledge, experience, abilities, attitudes, which each person has in order to generate strategic knowledge into the organization. On the other hand, it is possible to find a relational capital (Rc) which refers to the knowledge of the business environment such as customers, suppliers, users, beneficiaries, agreements and strategic alliances, the market, and the needs of the environment. Similarly, the structural capital (Stc) corresponds to the databases, operating manuals and information systems which the Organization must have, among others [15].

4 Model of Intellectual Capital and Its Relations with Other Types of Capital

The Knowledge Management (KM) has aspects productive, competitive and accounting, aspects that are reflected in its Intellectual Capital (IC). This capital is not only financial, but that becomes a tool used in the business environment and academic in the modern organizations. The models of KM university are supported by research processes that are supported by the human capital of the higher education institutions. For the development of the intellectual capital, it is required to promote and to participate in the research communities, with an important relational capital that integrates the different sciences of knowledge environment to the social development axes [16].

In addition, the knowledge management (KM) strengthens processes to use, share and develop the knowledge that the individuals and research groups possess, from their needs and their experiences supported by documents, articles, prototypes, methodologies and databases that can help to achieve the purposes of the university organization. Taking this point of view, it is necessary to create a knowledge management system that allows to build progressively a procedure based on the needs, applications of knowledge and the experiences of the research groups [17].

The intellectual capital (IC) is the main source of heritage in the organizations which possess the ability to identify, auditing, measuring, to renew, to enhance and to manage intangible assets, being a major factor to achieve the sustainable competitive advantages. At the same time, it makes to have successful organizations. For this reason, the managers focus their actions on the effective management of this capital. In this sense, researchers recognize the need of measuring and managing the intellectual capital, as a result the analysts and proponents propose some models of measuring and managing intangible resources [18].

Therefore, the university of KM models are based on the processes of research that support and consolidate the intellectual capital (IC), with the purpose of developing the processes required by the organizers and participants of research groups, as is the human capital (Hc), it also aims to generate a relational capital (Rc) and an institutional level that are equipped with a structural capital (Stc), according to their skills and knowledge, with which supports the social development of a country.

In the academic sector, it is presented a series of features in terms of these types of capital which includes: The human capital (Hc) whose elements are: teachers, students, academic and administrative personal. The relational capital (Rc) is composed by the social capital (Sc) and the business capital (Bc). The social capital contains certain elements such as: research and interaction, research policies, development of projects, publications, social relations, agreements with other institutions, university ranking, and other aspects of the social field.

The business capital, whose elements are: administration, organization and academic management, financial management, bodies and decision-making levels, university extension relations, relations of academic accreditation programs which are always focused on the processes of academic training, research and social interaction. On the other hand, the structural capital (Stc) is made up of the components of the technological and organizational capital [19].

The technological capital (Tc) includes an effort in the research, in the development of innovation, in the technology, in the intellectual and industrial property, in the results of innovation, in the educational resources, and in the organizational capital (Oc) that is formed by elements such as the legal and institutional standards and curriculum.

In the following diagram (Fig. 1) it is shown the relationship of intellectual capital with other types of capital and the internal and external environments. In addition, it must be taken into account that the elements, the variables and the indicators that depend on the specific academic organization.

Based on the previous ideas, the intellectual capital (IC) is formed by the human capital (Hc), relational capital (Rc) and structural capital (Stc). On the other hand, relational capital (Rc) includes other elements such as: the social capital (Sc) and the business capital (Bc); Besides this, it is included the structural capital (Stc) that contains other types of capital such as: organizational (Oc) and technology (Tc). These types of capital are developed in social, technological and organizational environments and extension. The interrelationship is given by training, by innovation and by the development of the intellectual capital supported by the management knowledge. In addition, it is intervened other actors such as: the academic, the administrative, the scientific, the social and the economics that are related by the information, knowledge and technological support flow.

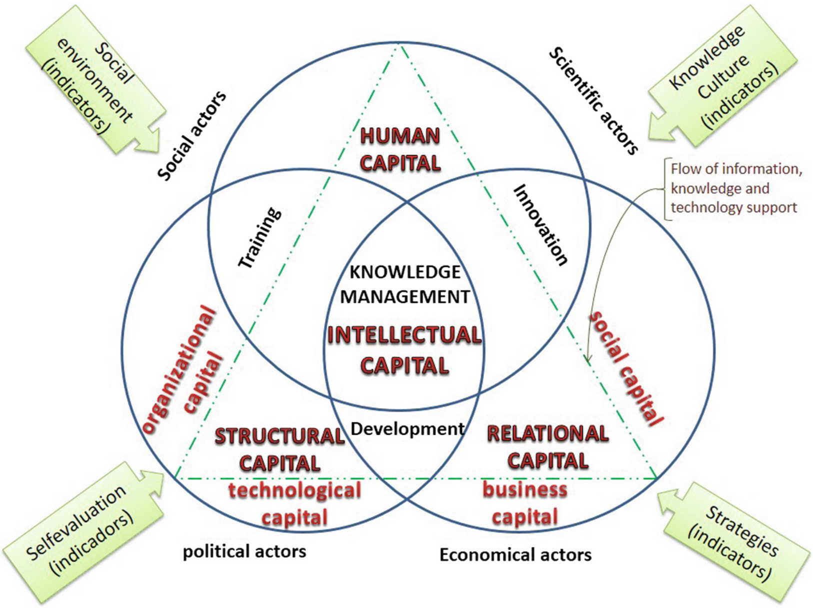


Fig. 1. Model of intellectual capital and its relations with other types of capital. source: authors.

5 Model of Intellectual Capital: Elements and Indicators

The model of intellectual capital is structured and organized by different types of capital. It contains a series of items which attach a set of variables, which are composed by a few percentage rates, as shown in followed Fig. 1:

This pentagon was created in order to facilitate the visual presentation and the important interrelationships of the different types of capital as the intellectual capital, in a university organization. With the goal of understanding comprehensively the various capital cities, and the factors that are converging, it is dealt with these elements: human capital (Hc) that contains the academic administrative staff element (AAE), teaching element (TE), university student element (USE), among others.

The relational capital (Rc) contains other types of capital as the social capital (Sc) with the elements of research and social interaction (ERSI), social relations, (SR), the university ranking (UR) and other; it also includes the business capital (Bc) with its academic, administration and management elements (AAME), related with the university students (RUS), in the institutions for promoting and improving the quality, and the accreditation of academic programs (IPIQA) (Fig. 2).

Finally, it is mentioned that the structural capital (Stc) contains other types of capital such as: organizational capital (Oc) with its legal institutional rules elements (LIRE), curriculum (CU), and others. Also it is possible to find the technological capital (Tc) with its elements: the intellectual and industrial property (IIP), the technology development (TD), the innovation results (IR), the educational resources (ER), the research effort (RE), the development, the research and the innovation (DRI), and others.

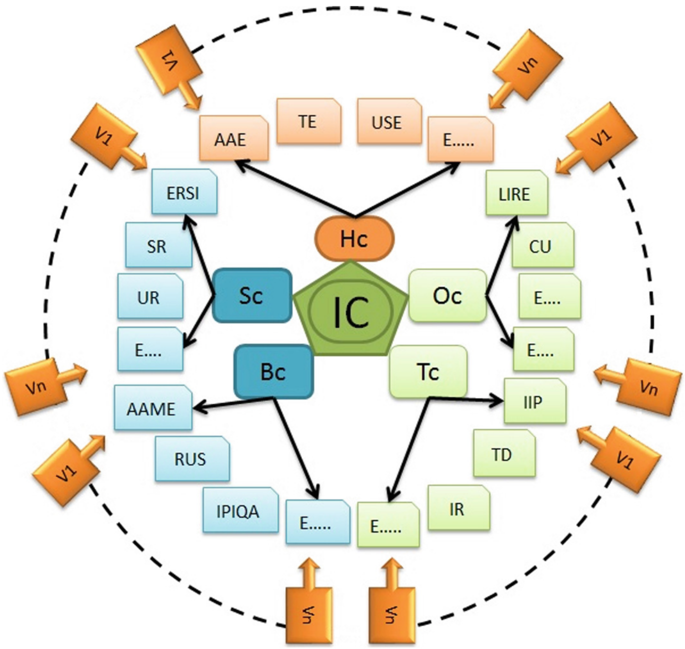


Fig. 2. Model of intellectual capital, its elements and variables. source: authors

Therefore, the intellectual capital is gotten by the standardized weighted sum of the different capitals. Each capital consists of the standardized weighted sum of the different elements. These elements contain standardized weighted variables that are measured by means of different indicators.

6 Mathematical Equation Proposal and Simulation

Based on the principles defined by considering the measurement model of the intellectual capital, the following equations are defined:

The intellectual capital is composed by the sum of the weight for every capital by using the following equation:

$$\mathrm{CI} = \sum\nolimits_{i=0}^{5} \mathrm{p}i\mathrm{C}i,\ \text{where} \sum pi = 1\ \text{and}\ \mathrm{Ci} = \text{capital's}. \tag{1}$$

Where C_i = Type of capital (human capital, organizational capital, technological capital, share and cardinal extension capital), p_i = is the weigh of every capital. The sum of the weights of each standardized weighted capital must be equal to one.

The types of capital are defined according to the sum of the weight by every element by using the following equation:

$$\mathrm{C}i = \sum\nolimits_{j=1}^{\mathrm{nj}} \alpha_j \mathrm{E}_j, \text{ where } \mathrm{n_j} \text{ number of elements } \mathrm{E_j} \text{ and } \sum\nolimits_{j=1}^{nj} \alpha_j = 1. \quad (2)$$

Where E_j = Type of element that belongs to every capital, α = weigh for every element, n_j is the quantity of elements that possesses every capital.

The elements of every capital are defined according to the sum of weight of every variable by means of the following equation:

$$\mathrm{E}j = \sum\nolimits_{k=1}^{nk} \mu_k \mathrm{V_k} \text{where } \mathrm{n_k} \text{ number of variables } \mathrm{V_k} \text{ and } \sum\nolimits_{j=1}^{nj} \mu_j = 1. \quad (3)$$

Where n_k = type of variable that belongs to every capital, μ = weigh for every variable, n_k is the quantity of variables that possess every element.

The variables of every element are defined according to the sum of weight of every indicator by means of the following equation:

$$\mathrm{V_k} = \sum\nolimits_{l=1}^{nl} \lambda_l I_l, \text{ where } \mathrm{n_l} \text{ number of indicators } \mathrm{I_l} \text{ and } \sum\nolimits_{l=1}^{nl} \lambda_l = 1. \quad (4)$$

Where n_l = type of percentage indicator, which belongs to every capital, λ = weigh for every indicator, n_l is the quantity of indicators that every variable possesses.

Later, a survey simulation was designed with few percentage indicators with the main purposes of measuring the presented variables; besides, with these variables, a series of elements was calculated; then, with the obtained elements, the types of capital were calculated; the sum of the types of capital generates the intellectual capital.

It is important to clarify that all the information of the indicators (I) are taken in percentage terms.

In this way, the percentages for every component of the model can be calculated by estimation of linear models generalized for analyzing the main components (ACP), looking for indicators from associations of covariance, between the values of the indicators by reducing its components, in order to find the percentage of final intellectual capital (latent variable) in a university organization. On the other hand, in the future, possible indicators will be defined accordance with every capital an example as it is appeared in the following Table 2:

As an example in the human capital, one of its elements is teachers; these teachers have or do not graduated degree that can be measured as a percentage of teachers with postgraduate studies. Another example is the technological capital, one of its elements are educational resources that can be measured as a percentage of equipment by a classroom.

7 Analysis of the Model of the Intellectual Capital Through Its Principal Components

The analysis of the major components (ACP) is given by the statistical problems that have a large number of variables from where you can choose the factors that best interpret the model. Within the multivariable techniques known as factorial methods, this analysis is a statistical technique that searches to synthesize a set of variables by

Table 2. Indicators of the different types of capital. source. authors.

Indicators of human capital	Percentage of teachers with postgraduated studies	Ich1
	Percentage of teachers for student	Ich2
	Percentage of teachers for program	Ich3
	Percentage of teachers plants	Ich4
	Percentage of teachers for contract	Ich5
	Percentage of gone away from pregrade	Ich6
Indicators of social capital	Percentage of projects for investigation lines	Ics1
	Percentage of projects of grade for program	Ics2
	Percentage of agreements inter - universities	Ics3
	Percentage of the lines of investigation for program	Ics4
	Percentage of national and international congresses	Ics5
	Percentage of teaching members of international scientific societies	Ics6
Indicators of business capital	Percentage of national and international conventions	Icn1
	Percentage of academic programs with accreditation	Icn2
	Percentage of academic programs with accreditation of quality high	Icn3
	Percentage of graduates follow-program	Icn4
	Percentage of annual evaluations of the academicians' organisms	Icn5
	Percentage of annual evaluations of the organisms of administration	Icn6
Indicators of organizational capital	Percentage of compliance of the institutional development plan compliance	Ico1
	Percentage regime of regulation of the student 's	Ico2
	Percentage of compliance plan for faculty development compliance	Ico3
	Percentage of objectives of the plan of studies	Ico4
	Percentage of use of laboratory equipment and audiovisual	Ico5
	Percentage of compliance with regulation of the teaching system	Ico6
Indicators of technological capital	Percentage of computer equipment in laboratories	Ict1
	Percentage of equipment by classroom	Ict2
	Percentage of access to the Internet	Ict3
	Percentage of offices and service areas by teaching general	Ict4
	Percentage of Software used with license	Ict5
	Percentage of specialized software used with license	Ict6

creating new variables that are linear functions which seeks to reduce the dimension of a set of number of variables in order to improve its interpretation. It is used because many variables produce too many coefficients which have a strong correlation between them.

Its methodology is done through faces which are explained below: Phase 1, analysis of the matrix of correlations: when large number correlations are presented and the chosen factors will explain the total variability. Phase 2, Selection of factors: the first factor explains the greater number of variables, the second, the maximum variability is not explained by the first factor, and so on. It takes into account those factors that explain the higher percentage of variability. These factors are called principal components. Phase 3, the analysis of factorial matrix: it represents the main components in the form of an array, each element will be the coefficients of the factorial variables, the array will be columns (CP) and rows (variables). Phase 4, interpretation of the factors: in order to make that a factor can be interpreted easily, you must have the following characteristics: factorial coefficients near 1 or -1, the variable that must have a single coefficient factorial high only with a factor which does not have similar coefficients factorials. For our example, according to the initial data that percentage are indicators, their correlation is standardized indicators and it was determined in the models for each capital as follows: According to the initial data, the highest correlation level for each capital by each indicator is presented as follows: for the human capital is explained in a 63.81 %, to the social capital is explained in a 61.58 %, for the business capital is explained in a 71.28 %, for the organizational capital is explained in a 62.62 %. Finally, for the technological capital, it is explained in a 60.23 %. Although, the average correlations of the indicators is not high enough considered that explanation about all models by the capital, it is explained a 63.90 %.

According to the previous explanation, the highest correlation for each capital by each indicator is presented in the following Table 3:

The weighted indexes by capital of each major component: based on the above indicators, it can be concluded that are more significant those whose values are closer to the absolute value of 1 or have coefficients factorials near 1 or -1.

Finally, the equations by capital according to their level of correlation can be expressed in this way:

$$\begin{aligned} \mathrm{Hc}:\ & 0.704 * \mathrm{Ich1} + 0.347 * \mathrm{Ich2} + 0.314 * \mathrm{Ich3} \\ & + 0.870 * \mathrm{Ich4} + 0.004 * \mathrm{Ich5} + 0.739 * \mathrm{Ich6} \end{aligned} \tag{5}$$

$$\begin{aligned} \mathrm{Sc}:\ & 0.613 * \mathrm{Ics1} + 0.725 * \mathrm{Ics2} + 0.836 * \mathrm{Ics3} \\ & - 0.570 * \mathrm{Ics4} + 0.706 * \mathrm{Ics5} + 0.118 * \mathrm{Ics6} \end{aligned} \tag{6}$$

$$\begin{aligned} \mathrm{Bc}:\ & 0.621 * \mathrm{Icn1} + 0.710 * \mathrm{Icn2} + 0.717 * \mathrm{Icn3} \\ & - 0.406 * \mathrm{Icn4} - 0.912 * \mathrm{Icn5} - 0.233 * \mathrm{Icn6} \end{aligned} \tag{7}$$

$$\begin{aligned} \mathrm{Oc}:\ & 0.928 * \mathrm{Ico1} + 0.654 * \mathrm{Ico2} - 0.108 * \mathrm{Ico3} \\ & - 0.145 * \mathrm{Ico4} + 0.830 * \mathrm{Ico5} + 0.147 * \mathrm{Ico6} \end{aligned} \tag{8}$$

$$\begin{aligned} \mathrm{Tc}:\ & 0.232 * \mathrm{Ich1} - 0.913 * \mathrm{Ich2} - 0.471 * \mathrm{Ich3} \\ & + 0.783 * \mathrm{Ich4} + 0.430 * \mathrm{Ich5} + 0.374 * \mathrm{Ich6} \end{aligned} \tag{9}$$

.

Table 3. Indices weighted capital of each main component. source: authors.

Hc	%	Sc	%	Bc	%	Oc	%	Tc	%
Ich1	0.704	Ics1	0.613	Icn1	0.621	Ico1	0.928	Ict1	0.232
Ich2	0.347	Ics2	0.725	Icn2	0.710	Ico2	0.654	Ict2	−0.913
Ich3	0.314	Ics3	0.836	Icn3	0.717	Ico3	−0.108	Ict3	−0.471
Ich4	0.870	Ics4	−0.570	Icn4	−0.406	Ico4	−0.145	Ict4	0.783
Ich5	0.004	Ics5	0.706	Icn5	−0.912	Ico5	0.830	Ict5	0.430
Ich6	0.739	Ics6	0.118	Icn6	−0.233	Ico6	0.147	Ict6	0.374

Where the extension the capital and social capital are more representative than the technological capital and would be focalized by investment and development.

According to the results for the intellectual capital, it is more significant the teachers with postgraduate studies and the increase of the teachers of plant, tan the percentage of teams of computing and the licensing of general programs, the fulfillment of the development plan for subject programs and the fulfillment of the regulation teachers, is more important than the national agreements and project graduation program; for the business capital the percentage of academic programs with accreditation and percentage of agreements with private enterprises has more significance than a set of indicators of agreements with the state institutions and percentage of evaluation of academic organisms.

It is necessary to bear in mind that the intellectual capital would be obtained of the weighted sum of the different types of capital.

It is clarified that the analysis of the major components (ACP) the percentages of the indicators are between 1 and minus 1, showing the degree of correlation between them, looking for the most relevant. Later, in order to find the intellectual capital, it must be weighted standardize weights of each capital which sum must be equal to one. As an example, the Eq. (5) is thus:

$$\begin{aligned} \mathrm{Hc:}\ & 0,236 * \mathrm{Ich1} + 0,117 * \mathrm{Ich2} + 0,105 * \mathrm{Ich3} \\ & + 0,292 * \mathrm{Ich4} + 0,001 * \mathrm{Ich5} + 0,248 * \mathrm{Ich6} \end{aligned} \tag{10}$$

.

8 Conclusions

Considering the ideas previously exposed, it can be stated that the knowledge management are part of the assets of a professional academic community. Similarly, it can be said that the intellectual capital consists of some basic components such as: the human capital, the relational capital and the structural capital which contain other types of capital variable elements and indicators that can be measured. In this way the knowledge management can be better interpreted.

Similarly, there are different models that have a high level of applicability in the various academic institutions, but it really depends on the special conditions, on which those organizations are developed.

The contribution of this article is that when measuring intellectual capital, you can make decisions on issues of educational policy and as a result the relationship between the university and society can be optimized. If we can measure it, it is possible to improve it.

At the same time, it is concluded that the intellectual capital can be measured and can be given a valuation of the intangible assets, which generates new knowledge in the universities, enterprises and society.

To integrate the intellectual capital to the knowledge management in the university it is needed to improve the different interrelated types of capital, with the participation of the social, economic, administrative and scientific actors in the society with which it hopes to achieve the development of the society.

At the university level, the researching directors are obliged to be organized in an interdisciplinary community, in order to reinforce their skills and enhancing their relations with the business and the society, because it will optimize the knowledge management in different areas.

On the other hand, it is clear that the university does not give enough importance to the intellectual capital for this reason probably, it is not considered as a tangible asset, although it is suggested to consider the intellectual capital as a potential advantage that can offer monetary benefit for the academic institutions.

It has been explained some intellectual capital models, but its analysis are based on linear models and non-weighted, which do not reflect the interrelationship and weight that can have the different elements that integrate the capital, those variables and indicators could be the value of intellectual capital, and thus it would allow the possibility of incorporating them, as an asset for a university organization.

Finally, the model of the intellectual capital with its components and their weights can be calculated by estimating or using principal components analyses, being focused on those that are more significant and better explained for the behavior of their types of capital in a university organization.

The principal components analysis applied when you have multiple variables, highly correlated would let you to avoid redundancy: in this sense, it would be possible to find the most significant variables within the model. The implementation of the proposed model will depend on the information of each university organization.

References

1. Fundibeq. La gestión del conocimiento. (2014). http://www.fundibeq.org/opencms/export/sites/default/PWF/downloads/gallery/methodology/tools/gestion_del_conocimiento.pdf
2. Muhammed, S., Doll, W.J., Deng, X.: Developing individual level outcome measures in the context of knowledge management success. strategies for knowledge management success. Hershey: Information Science Reference, p. 123, USA (2011). doi:10.4018/978-1-60566-709-6.ch007
3. Bueno, C.E.: Gestión del Conocimiento en Universidades y Organismos Públicos de Investigación (2001). http://www.madrimasd.org/informacionidi/biblioteca/publicacion/doc/16_GestionConocimientoUniversida-desOPIS.pdf. Accessed 18 January 2008
4. Clavijo, M.: Cómo retener el talento. Journal Gerente, No. 156, primer semestre (2011)

5. David, P., Foray, D.: Fundamentos económicos de la sociedad del conocimiento. J. Comercio Exter. **52**(6), 472–490 (2002)
6. González, J.: Aplicación y desarrollo de la Gestión del Conocimiento de los grupos de investigación de la UPTC. Universidad Nacional de Colombia, Bogotá (2009)
7. Moreno, L.: Región y sociedad. Los negocios en la era digital. vol. 1. México: Plaza & Janés (2002)
8. Plaz, R. Gestión del Conocimiento: Una visión integradora del aprendizaje organizacional. Recuperado de (2003). http://www.madrimasd.org/revista/revista18/tribuna/tribuna2.asp
9. Zorrilla, H. La Gerencia del Conocimiento y la Gestión Tecnológica. Programa de Gestión Tecnológica (1997). http://www.sht.com.ar/archivo/Management/conocimiento.htm
10. Clemmons, M.: Knowledge Management, Exploring the Oxymoron, vol. 1. Alpha Books, USA (2002)
11. García, I.: La gestión del conocimiento el nuevo paradigma de las empresas (2002). http://cdt.robotiker.es/cdt/articulos/mostrar_pdf.jsp?id=1
12. Arbonies, Ángel Conocimiento para Innovar. Vol. 1, 2a. Ed. Madrid: Díaz de Santos (2006)
13. Medina, V., Pérez, N., Torres, J.: La Investigación en Ingeniería: Ciencia y Tecnología de la Información y del Conocimiento. Bogotá: Fondo de Publicaciones de la Universidad Distrital (2011)
14. Euroforum. Medición del Capital Intelectual. Modelo Intelect, Madrid, Euroforum (2014). http://web.archive.org/web/20120428095005 http://www.gestiondelconocimiento.com/index.php3, Acceso: 25 de Agosto del 2014
15. Huanca, J. Modelo de gestión de conocimiento para la gestión académica de las universidades públicas de Bolivia (2012). http://www.virtualeduca.info/ponencias2013/87/JuanCarlosHuancaModelodeGestiondeConocimientoenlaUniversidad.doc
16. Rodríguez, M., González, J.: Gestión del Conocimiento y Capital Intelectual, a través de modelos universitarios. Revista Económicas CUC **34**(1), 85–116 (2013). Editorial Educosta, Barranquilla
17. Guevara, J., Lara, J., Moque, C.: Sistema de gestión de conocimiento para apoyar el trabajo de grupos de investigación. J. Tecnura **16**(33), 97 (2012). julio-septiembre de
18. Carrillo, E., Gutiérrez, F., Díaz, C.: Propuesta de indicadores para gestión del capital estructural en grupos de investigación. Universidad & Empresa No. **22**, 102 (2012)
19. Díaz, F.N., Medina, V.H., y González, R.: ICT as a means of generating knowledge for Project Management. In: Uden, L. et al. (eds.) The 8th International Conference on Knowledge Management in Organizations. National University of Kaohsiung, Kaohsiung, Taiwan. Springer Proccedings in Complexity, 9–13 September 2013, pp. 617-629, ISSN: 2213-8684, ©Springer + Business Media Dordrecht. 2014 (2013). doi:10.1007/978-94-007-7287-8_50

Social Capital-Knowledge Management (SCKM) Analysis Framework on an E-Portfolio

Wardah Zainal Abidin[1(✉)], Lorna Uden[2], and Rose Alinda Alias[3]

[1] Department of Informatics, Advanced Informatics School, Universiti Teknologi Malaysia, 54100 Kuala Lumpur, Malaysia
wardah@utm.my
[2] FCET, Staffordshire University, The Octagon, Beaconside, ST18 OAD Stafford, UK
l.uden@staffs.ac.uk
[3] Office of Deputy Vice Chancellor (Academic and International), Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor Darul Ta'zim, Malaysia
alinda@utm.my

Abstract. The introduction of ePortfolio by most universities circa 2006 was to respond to stakeholders claims and demands that graduates were lacking in generic skills. Hence the main objective of the usage of eP among undergraduates was to document their acquirement of generic skills. One such project in Malaysia however did not achieve much success. The authors believe that the process of acquiring generic skills by students is the fundamental step for a student to build his/her very own social capital. This paper attempts to examine the failed system from the perspective of Social Capital theory and knowledge management (SCKM). The examination considered both the dimensions of social capital and the processes of knowledge management. The analysis revealed that the system lacked some fundamental elements of social capital in its development as well as during its roll-out. This leads the paper to offer some design and implementation suggestions for other similar undertakings.

Keywords: Social capital · Knowledge management · ePortfolio · Social capital-knowledge management analysis framework

1 Introduction

The introduction of electronic portfolios (ePortfolio) in higher institutions was among the many innovations which universities around the world had to embrace as a response to a global outcry from the industry stakeholders circa 2006. The lack of employability skills portrayed by fresh graduates was among their major concern and has since pushed universities to embark on a more student-centred education. With students at the centre of the teaching and learning cycle, the use of student e-portfolios to archive their learning experience seemed a logical platform to elevate their self confidence.

Self confidence comes from the many and varied relationships which students build among their friends and outside their circles. The more the relationships which the

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 754–769, 2015.
DOI: 10.1007/978-3-319-21009-4_55

students have made the more they have gained in their social capital status. Hence the use of ePortfolios by students to showcase their social capital to interested parties is indeed a competitive advantage among universities. However not many studies have shown the direct link between ePortfolios and social capital among university students. This paper attempts to show this link via a case study method by examining an ePortfolio system from the perspective of both Social Capital theory and knowledge management theory.

This paper begins with related works on ePortfolio, knowledge management and social capital. A brief write up on case study as a research methodology precedes the section on analysis of a case study on the shortcomings of a campaign by a Malaysian public university to popularise its ePortfolio system among its undergraduates (Zainal Abidin 2008). This analysis utilizes the social capital-knowledge management (SCKM) framework which highlights the shortfalls of the ePortfolio system from the perspectives of SCKM dimensions. It concludes with suggestions for an enhanced design and implementation propositions of an ePortfolio system as well as future works.

2 Related Works

2.1 Electronic Portfolio

EPortfolios support a holistic approach to learning. Using ePortfolio helps students in their learning beyond simply attaining a grade in a course by thinking about how to clearly articulate what they have learned as well as to identify areas for improvement or further learning (Tosh et al. 2005). This means that students have to reflect on their own learning and by doing so, will be able to better integrate their various learning experiences.

According to the American National Learning Infrastructure Initiative (NLII), an ePortfolio is “a **collection** of authentic and diverse evidence, drawn from a larger **archive** representing what a person or organization has learned over time on which the person or organization has **reflected**, and designed for **presentation** to one or more audiences for a particular rhetorical purpose” cited in (Barrett, 2005). It consists of two items–the digital archive and the presentation drawn from it. It must be seen both as process and product.

The above NLII definition focuses on contents, built as the result of a learning and reflection process. Serge Ravet (2005) argues that ePortfolios are also the foundation for the provision of many different services, beyond presentation, for example:

- Supporting learning activities
- Accreditation of Prior Learning
- Managing continuing professional development (CPD) and personal development planning (PDP)
- Sharing knowledge within a community or an organisation
- Reflecting on one’s learning and/or practice
- Course registration

Eynon et al. (2014) proposes that (1) ePortfolio initiatives advance student success; (2) making student learning visible, ePortfolio initiatives support reflection, social pedagogy, and deep learning; and (3) ePortfolio initiatives catalyze learning-centered institutional change.

An ePortfolio can also be viewed as a digital profile, a digital representation used to support interaction between the ePortfolio owner and other individuals, communities or organisations (Ravet 2005). An ePortfolio can be created and used by individuals, communities and/or organizations to (1) archive and share learning and culture acquired from informal, non–formal and formal learning environments; (2) provide evidence or verification of learning in a dynamic, multimedia fashion; (3) manage personal and collective learning to maximize usage and to plan; and (4) showcase or promote particular assets as required (Chang et al. 2011).

According to (Chen and Black 2010), an ePortfolio captures and documents the students' learning, reflection, rationale building, and planning, thereby establishing a culture that shares personal learning and promotes student-centered learning.

EPorfolios can be used to encourage further peer collaboration through the sharing of work, and promote analysis and reflection – enhancing learning and knowledge sharing. According to (Banks 2004; JISC 2008), another benefit of an ePortfolio is in the context of the workplace and lifelong learning which relate to the development and utilization of social capital. Social capital can enhance an organization's ability to manage knowledge because it has the capacity to do a variety of things.

Interestingly there is a dearth in the literature depicting the relationship between social capital and ePortfolio among students while the latter is also considered one of the many KM tools (Zainal-Abidin et al. 2014). Students, especially undergraduates will benefit most with the assembling of their social capital starting from their university days. And one natural platform to do this is, is via the ePortfolio.

2.2 Knowledge Management (KM)

Knowledge forms the basis for sustained competitive advantage (Teece 1998). The number of organizations claiming to work with KM has since grown progressively (Davenport and Guest 2001). According to (Turban et al. 2008), knowledge management is defined as the process of accumulating and creating knowledge, and facilitating the sharing of knowledge so that it can be applied effectively throughout the organization. Knowledge management involves four main processes. The first process is the generation of knowledge, which includes all activities that discover "new" knowledge. The second process is knowledge capture, which involves continuous scanning, organizing, and packaging of knowledge after it has been generated. Knowledge codification is the third process and it is the representation of knowledge in a manner that can easily be accessed and transferred. The fourth process, knowledge transfer, involves transmitting knowledge from one person or group to another person or group, and the absorption of that knowledge (Pearlson and Saunders 2004).

The importance of knowledge lies in the creativity value that it adds to the organization's assets, and in its ability to improve the effectiveness of an organization's intellectual capital (Sullivan 1999). Knowledge management can help managers to improve their day-to-day work, decision-making processes, create new responses, and enable a set of competitive reactions to be augmented (Belaid Kridan and Steven Goulding 2006).

Knowledge represents a key source for sustained competitive advantage according to (Drucker and Drucker 1994) and it is only successful if it has a strategic orientation approach to manage its stock of knowledge (Edvinsson 1997). It is important for organizations to know what factors determine an organization's capacity for knowledge management capabilities. Among these factors, social contexts of organization have critical impacts. The relevance of social capital (SC) for knowledge management (KM) has been discussed by researchers (Jaview Carrillo et al. 2006; Manning 2010; Smedlund 2008).

2.3 Social Capital

It is no longer sufficient to compete on content in a knowledge economy. Everyone has access to a multitude of content. Because everyone is highly skilled and experienced at the top, it is hard to compete on individual competency when everyone is so similar. According to Krebs (2008), the new advantage is context – how internal and external content is interpreted, combined, made sense of, and converted to new products and services. He further argues that creating competitive context requires social capital – the ability to find, utilize and combine the skills, knowledge and experience of others, inside and outside of the organization. Social capital is derived from employees' professional and business networks. Social capital is what connects various forms of human capital.

Today we have software to support the building and tracking of social capital within and between organizations. The two most popular social networking services are LinkedIn and Facebook. Both allow employees to track their business or social networks and provide a place online for people to meet and keep up with their associates, colleagues and friends.

The authors concur with (Krebs 2008) that social networking sites have some problems. First, they are outside the firewall of the corporation. Detailed employee career and contact information should not be shared on these sites. Second, these sites are patrolled by head hunters and your competitors. You can't hide on the Internet, but employees (and HR) should be careful what corporate information and structures are shared on these public social networking sites.

Last, but not least, the business and professional networks on these sites can be very inaccurate (Krebs 2008). An employee connected to innovators and thought leaders in his or her field has valuable social capital to do their job, and share with their corporate colleagues. An employee that is connected to many others, who are selfishly interested only in their own transactions, does not have social capital that is useful to their employer. It is our belief that ePortfolio serves as a powerful tool to be used as social capital for knowledge management.

According to (Monavvarian et al. 2013; Jaview Carrillo et al. 2006) social capital consists of knowledge and organizational resources that enhance the potential for individual and collective action in human social systems. Social capital comprises of those resources that actors may access through social ties that may affect an individual's action directed toward another based on the social structure in which the action is embedded and the history of transactions between the actors Monavvarian et al. (2013).

It is "the sum of the resources, actual or virtual, that accrue to an individual or group by virtue of possessing a durable network of more or less institutionalized relationships of mutual acquaintance and recognition" (Bourdieu and Wacquant 1992). It represents resources that reside in function-specific social relationships in which individuals are embedded. It is suggested by researchers that organizations with high levels of social capital have more knowledge-management capabilities than organizations with low levels of social capital (Hoffman et al. 2005).

Lang (2004) suggested that Social capital serves three important functions. Firstly, it represents a structure of obligations, expectations, and trustworthiness. Secondly, it serves as information channels. Finally, social capital serves as a system of norms and effective sanctions, that is, effective norms which constrain people from socially undesirable actions. Hoffman et al. (2005) suggested that social capital can be separated into five distinct dimensions. They are information channels, social norms, identity, obligations and expectations, and moral infrastructure (Table 1).

Five Dimensions Together. Although each of the five dimensions of social capital is separate and each provides distinct benefits to the organization, they are mutually dependent on each other for their development. For example, without strong information channels that create strong ties between individuals within the organizational network, there is no opportunity for the organization to experience closure (Coleman

Table 1. Dimensions of social capital

Dimensions	Summarised explanation
Information channels	- Are Social Networks within the organization; as mechanisms that connect to the outside world; most obvious inventory of social capital; have value (Putnam 2001) - Also contain formal structure of an organisation; consists of inter-personal relationships developed over the course of relationship. Provides major benefits (are abundant and strong ties within the network) which provide closure (Coleman 1988) to guarantee observance of social norms - Are important form of SC in providing contemporary and contextualized information that are essential for KM, Manning (2010); facilitating the process that permits the development of tacit, experiential knowledge in 'learning the ropes'; corresponds closely to the skills-based, difficult-to-codify, insider knowledge
Social norms	Definition: norm is a norm is a property of a social system, not of an actor within it (Coleman 1988); specify what actions are regarded by a set of persons as proper or correct, or improper or incorrect - provide for social control in an organization; general, internalized sets of accepted behavior for members of the social network - are shared strategic visions, systems of meanings, and normative value orientations (Nahapiet and Ghoshal 1998) - are a common belief system that allows participants to communicate their ideas and make sense of common experiences (Adler and Kwon 2000) - are also the accumulated history of the organization in the form of social structure appropriable for productive use by any member of the

(*Continued*)

Table 1. (*Continued*)

Dimensions	Summarised explanation
	social network in the pursuit of his or her interests (Sandefur and Laumann 1998)
Obligations & Expectations	- are the positive interactions that occur between individuals in a network, Lesser (2000); positive largely because of the levels of trust and reciprocity that they engendered (Putnam 1993) - provide future benefits which are nurtured in an organizational environment containing strong social ties and are hampered by the absence of these ties (Hoffman et al. 2005) - can lead to networked collective trust, which becomes a potent form of expectational asset (Nahapiet and Ghoshal 1998) - are strengthened by collective trust (Hoffman et al. 2005), where group members can rely on one another to follow through with things expected of them and obligations owed by them; are then more willing to work for the group with the knowledge and expectation that the group will work for them when their time comes - allows group members to rely on each other more generally to help solve the everyday problems of cooperation and coordination (Hoffman et al. 2005)
Identity	- occurs when individuals see themselves as one with another person or group of people (Nahapiet and Ghoshal 1998); Individual takes the values or standards of other individuals or groups as a comparative frame of reference (Merton 1968) - Group identity increases perceived opportunities for information exchange and enhances frequency of cooperation (Lewick and Bunker 1996) - conversely if not present there are significant barriers to information sharing, learning, and knowledge creation (Child and Rodrigues 1996)
Moral infrastructure	- is identified as the structure or network, which allows an organization to encourage norms of conduct within the organization's scope of influence, Hoffman et al. (2005); - are networks of civic engagement at the community level, Putnam (1993); Civic engagement refers to people's connections with the life of their community and includes such things as membership in neighbourhood associations, choral societies, or sports clubs (Blanchard and Horan 1998; Putnam 1993); - whether existent within an organization, or within a community, provide an additional pathway for network actors to learn of the trustworthiness of individual actors within the network; - provides additional closure for social norms and gives individuals, acting in their own rational self interest, solid reasons to act in ways that adhere to formal and informal codes of conduct in their organization (Blanchard and Horan 1998)

1988). Without closure there is no opportunity for the organization to develop strong social norms and for identity to begin to take hold. Without strong social norms there is no opportunity to develop a system of obligations and expectations and to provide for the adherence to a set of ethics, both formal and informal (the moral infrastructure). In many ways social capital at its core, is about the value created by fostering connections between organizational members. Among various aspects of social capital, trust has the most significant effect on the knowledge management effectiveness. This finding is consistent with prior researches done by (Putnam 1993) and (Fukuyama 1995) who stress the importance of trust in social relations. Another aspect is that social capital promotes group interaction and relation among members. Thus improving the quality of social capital in ePortfolio usage will foster better knowledge creation and knowledge sharing in the university.

How Social Capital can Enhance Knowledge Management within an Organization. According to (Hoffman et al. 2005), social capital can enhance an organization's ability to manage knowledge. In terms of knowledge creation, social capital helps to facilitate the development of collective intellectual capital by affecting the conditions necessary for exchange and combination to occur. Social capital can also facilitate the development of intellectual capital. Since intellectual capital depends on the combination of knowledge and experience of different parties, intellectual capital's creation is greatly facilitated by the existence of social capital. Social capital has also been theorized to play a role in the development of core competencies (Kogut and Zander 1996) that are vital to knowledge creation.

Social capital also increases the efficiency of action (Lesser 2000) and encourages cooperative behavior (Coleman 1988; Nahapiet and Ghoshal 1998). Additionally, social capital has been theorized to serve as an important element in the development of human capital (Coleman 1988) and to provide access to resources through network ties (Burt 1992). Social capital can also enhance knowledge capture, knowledge codification, and knowledge transfer. It enhances these knowledge management processes because it contributes to a firm's ability to create value in the form of innovation through the facilitation of combination and exchange of resources in a firm (Kogut and Zander, 1993).

According to (Fukuyama 1995) social capital can enhance the entire knowledge management process because it makes working together as a collective group more efficient, and can work without the need for formal contracts, incentives, and monitoring mechanisms that are necessary in systems with little or no social capital among organizational members.

Hoffman et al. (2005) argued that social capital is a determinant of an organization's capacity for knowledge management. Organizations with high levels of social capital have more knowledge management capabilities than organizations with low levels of social capital. This means that it has practical implications for e portfolio use and specifically for trying to enhance owners of ePortfolios' ability to manage knowledge in their learning.

It is our belief that SC is supportive of knowledge management. SC can foster creative, innovative, and initiative-taking behaviors among participants and sharing knowledge among individuals and groups in organization – i.e. behaviours that are linked to advantageous knowledge practices.

Trust is an essential component of social capital (Putnam 1993). He argues that trust facilitates cooperation. The greater level of trust within a community, the greater the likelihood of cooperation. Cooperation breeds trust. Fukuyama (1995) defines trust as the expectation that arises within a community of regular, honest, and cooperative behaviour, based on commonly shared norms, on the part of other members of that community.

3 Research Methodology

3.1 Case Study Method

In line with the criteria which suit the use of a case study method, this research which involves a Malaysian public research university satisfies all the three conditions. Firstly, the kind of research question which it is set to find out is either *descriptive* or *explanatory* in nature. In this case, the question which puzzles the management is, "Why is the ePortfolio not attractive to undergraduate students even though studies show that keeping and building ePortfolio is good for their success?"

Secondly, the study of a phenomenon such as this, must be within its real-world context, and data collected should be in its natural settings. While one may use a lengthy questionnaire to get the responses, however observing and doing original fieldwork can contribute effectively into understanding the phenomenon. Finally, many use the case study method to conduct evaluations.

In this research, the first author is directly involved in the university ePortfolio Task Force due to her vast experience working together with colleges under the ministry of human resource in the areas of skills development where the use of portfolio is fundamental and mandatory as authentic assessment tools. Hence data for this case study are derived mostly from seven years of participating as team members, observation, documents analysis, interviews, meeting minutes and sampling surveys on students.

3.2 Data Analysis Using SCKM Analysis Framework

To evaluate the potential usefulness of the SCKM analysis framework, data from the case study are scrutinised and dissected in terms of their functionality or characteristics from the SC and KM perspectives concurrently. This resulted in a 5 × 4 matrix. This will be dealt with in detail in the next section. Due to the limitation of the book size and layout, the 5 × 4 matrix is depicted by five separate tables.

4 E Portfolio Case Study

4.1 Background of the Case Study

About the middle of 2000, universities globally were opened to the idea of using portfolio and particularly ePortfolio to their undergraduates. This was among the actions taken by most universities around the globe partly as a response to claims made by the industry that university graduates were mostly lacking in employability skills to face the real world of work. Around the year 2007, a Malaysian public university which is heavily focused in engineering and technology, made a decision to embark on a university-wide use of ePortfolio for its undergraduates. Table 2 shows the facts about the first version of the ePortfolio developed by the university.

To examine the ePortfolio system built by the university in terms of its capability to support KM via the incorporation of social capital features, the authors have into Table 3. In Table 3 the horizontal axis features the KM processes (Generate, Capture, Codify and Transfer of knowledge) while the vertical axis features the five dimensions of SC (Information channels, Social Norms, Obligations & Expectations, Identity and Moral Infrastructure). The contents in the individual cells are abstractions of the ePortfolio system in terms of the different processes of managing knowledge. The left side of the table under SC outlines the ePortfolio system capabilities with regards to the five dimensions of social capital development.

Table 2. Salient facts about the university ePortfolio system

Facts about the university and the Implementation of ePortfolio
- 10 faculties; 4 are engineering faculties
- close to 1,700 first year students as fulltime and new users
- mandatory to first year, first semester students and later to continue using it until they graduate
- every student was given a 3-h. hands-on training in the first 2 weeks of the semester
- current senior students were encouraged but no formal training was given to them
- each student is assigned to one academic advisor mainly for supporting academic administrative matters
- the roles and responsibilities of the academic advisors are not changed from the pre-eP days hence not really suited to help students in developing and nurturing eP of students under their charge
- training to academic advisors in charge of 1st year students on guiding students with eP is given but not deep enough and is usually the last module (about 1 h) of a 2-day training session; usually the attendance of the participants is at its lowest point
- circulars and notices on the eP implementation are disseminated to all faculties in care of the Deputy Deans of Academic, who are already overwhelmed with the heavy demands of the newly implemented OBE (outcome-based education)
- OBE was just introduced with much opposition from the teaching fraternity around 2003-4
- although the raison d'être of the eP project was OBE-related i.e. to assist faculty deans with the aspects of evaluation of generic skills among students, this is however not well translated into fruition
- the eP was only meant for undergraduates (UG) while the contributory factors towards giving marks in achieving Research University (RU) status were not coming from UG but largely from Postgraduate-related activities like doing research, writing publications and supervising PG students
- random survey on the existence of eP to members of the university revealed that many are not aware of its existence, its use as well as its function

(*Continued*)

Table 2. (*Continued*)

Facts about the university and the Implementation of ePortfolio
Facts about the ePortfolio system
- initiated in 2006, by the University central unit called the Centre for Teaching and Learning (CTL) which is responsible for the training and advancement activities of all teaching and learning to lecturers - representatives of all faculties with UG programmes (deputy deans and IT managers) were called to participate in the initial 2-day retreat workshop to design the ePortfolio - the underlying premise of the ePortfolio system is to support the university especially for engineering programmes to be accredited and to show that these programmes comply to the teaching requirements of OBE - the main features were heavy towards producing the necessary statistics and reporting on the acquirement of generic skills by the students with regards to compliance to the PLO (programme learning outcome) of individual UG programmes - the in-house design specifications were then given to the university's centre for ICT (CICT) which was already overloaded with other in-house development of academic and administrative systems of the university - the current style of system development of the day is menu-driven real-time online interface with SDLC approach and available via internet to end-users - the programming language used is Apache-based PHP and the back-end database is Oracle 9 - the system is able to handle text-editing, table formatting and input of audio and video format with a maximum size of 2 GB per person - it is standalone and not linked to any other systems such as the current Moodle-based elearning system - the academic administration unit in each faculty is responsible to input (i) their respective PLOs into the system, (ii) list of academic advisors and students data which is quite cumbersome but recurring at the beginning of every new semester - the interface is template-based so that the ePortfolio owners (the students) can only input their data in the boxes provided - the available features are (i) about me, (ii) reflections against each PLO for every course taken in that semester, (iii) providing labels to every artefact input into the system and connecting it to a PLO, (iv) input of data to build a CV, and (v) comments from academic advisors - there is a time range for using the ePortfolio for a semester so that grading will be made by academic advisors after a cut-off deadline - all functions are available online and there is no off-line function - there is no feature allowing chatting between students or peers; but by providing the link anyone can see the ePortfolio of the individual owner

4.2 The SCKM Analysis Framework at Work

The 5 × 4 matrix is derived from the Social Capital's 5 dimensions namely, Information Channels, Social Norms, Obligations & Expectations, Identity and Moral Structure versus the KM's four processes which are Generate, Capture, Coding and Transfer.

The KM analysis will shed some issues in relation to the five dimensions of SC. In turn the breadth and depth of these dimensions populated as a consequence of the KM process will reveal the degree of Social Capital characteristics of the ePortfolio system. Five tables are built and these are Tables 3(i), (ii), (iii), (iv) and (v) as depicted below.

Table 3(i). Using SCKM: ePortfolio system as INFORMATION CHANNELS against KM processes

SC dimension: Information channels	KM processes			
	Generate	Capture	Coding	Transfer
- system was not too stable in accepting large multimedia files - system only accept pdf files for non-multimedia documents - the internet connectivity was not able to support a large number of users at one time - the internet accessibility was limited to public areas; not in hostels - no offline capability; totally internet and real-time system - system was drop-down menu interface and later changed to template structure - not linked to other student-related systems	- eP owners need to express their experiences via writing reflections; 1st year students find it a daunting task - faculty act as mentors provide advice and comments; most faculty are new to OBE style of teaching	- eP users need to scan text/image or record audio/video files prior to uploading and storing into eP system; facilities are only available in labs - CICT eP project team is undermanned; lack in providing integrated support services to serve students	- eP users need to know which artefacts goes into which PLO; 1st year students are not familiar with university teaching and learning jargons - most student activities are limited to academic - they are still in the searching mode to understand the different value systems of peers	- many levels of authorities for students to comply with - sharing experiential knowledge; most students are selective in sharing information with peers

From the above tables (Table 3(i), (ii), (iii), (iv) and (v)), it can be shown that the lack of having features to support the basic tenets of SC may suggest that the ePortfolio system was not that appealing to the main users of the system, i.e. the students. While the system scores high as information channels, it performs rather badly in the rest of the dimensions for social capital generation. The authors believe that by populating this 5 × 4 table, it can provide a more meaningful analysis on the ePortfolio system to show its performance as well as a means for future improvement on the system.

The above analysis therefore can provide an insight to the designers of the ePortfolio system to improve the system systematically in view of ensuring the system

Table 3(ii). Using SCKM: ePortfolio system as SOCIAL NORMS against KM processes

SC dimension: Social norms	KM processes			
	Generate	Capture	Coding	Transfer
- the concept of keeping a personalised portfolio is new and awkward among most students and lecturers (advisors) - the system is one-size-fits-all to cater for all levels of students and all types of academic programmes - limited feature of social-networking among users; no chat capabilities - lack features to store address book of counterparts and guests	- existing units in university interacting with the new students project different forms of social interactions - university management's attention is towards attaining Research University status and this is enforced to faculties	- lack understanding of social norms in the university - familiarity of school-type culture is still prevalent in the 1st semester for new students - depend greatly on group decision	- 1st year students inferiority complex towards seniors and faculty - categorisation of artefacts is not seen as an important activity	- "middle-up approach": initiated the centre for teaching and learning, a central unit; neither top-down nor bottom-up initiative

Table 3(iii). Using SCKM: ePortfolio system as OBLIGATIONS & EXPECTATIONS against KM processes

SC dimension: Obligations & Expectations	KM processes			
	Generate	Capture	Coding	Transfer
- system does not have prompts for students to write deep reflections - system is mandatory for first year first semester students who are inapt of expressing reflections due to communications limitations - system allows only advisors to give comments but no value in final grading - some advisors are not course tutors hence a gap in relationship value	- mandatory: 1st year undergraduate students - optional for senior students - grading of building eP is not part of transcript - students lack proper guidance from advisors - not all advisors are able to provide advise	- differing expectations of various stakeholders - lack of emphasis on its importance from top university management - lack of commitment from faculty - driven by extrinsic motivations only	- sense of unnecessary confusion to differentiate between academic and social expectations among 1st year students	- most academic advisors are already burdened with existing responsibilities

is not only supporting the university to document the needed evidences for generic skills acquirement among the students, but what is more important is the SC-friendly features of the system to assist students to start and accumulate their assets in networking, and career building.

Table 3(iv). Using SCKM: ePortfolio system as IDENTITY against KM processes

SC dimension: Identity	KM processes			
	Generate	Capture	Coding	Transfer
- menu-driven templates do not allow any form of creativity in organising their data - students are mere items of statistics in the university's quest to have evidence of generic skills acquirements against programme learning outcomes to show to accreditation auditors - only form of individuality is in writing of labels and comments for every artefacts which students download and share as their evidences	- sense of belonging to the university and to faculty is still budding (1st year students)	- intrinsic identity crisis of young adults fresh from school indoctrination - teamwork is paramount in getting things done; the new students lack the needed cohesive team structure	- advent and popularity of Facebook slowly overtaking Myspace - needs of 1st year students are different from those in final year	- projection of self-worth is not pronounced in most 1st year students

Table 3(v). Using SCKM: ePortfolio system as MORAL INFRASTRUCTURE against KM processes

SC dimension: Moral infrastructure	KM processes			
	Generate	Capture	Coding	Transfer
- no particular emphasis is given on this aspect - supports curriculum vitae generation of individual students but still conform to a set standard - value of having a good individual eP is not readily grasped and appreciated by 1st year students	- acculturation breeds over time - motivational factors for using eP are not intrinsic in the beginning; more of a chore and assignment	- 1st year students are still in culture shock stage in semester 1 - blurry of one's self achievement, hence important actions are not captured for evidence of generic skills acquirement	- fear of making mistakes - peer pressure to complete the eP is strong in the beginning of semester	- no seniors are familiar with the system for juniors to seek help from

5 Future Works and Conclusion

There are many potential benefits for using ePortfolio in learning and also knowledge sharing. Knowledge management is a complex and multi-dimensional process. Successful Knowledge management implementing in organization requires a comprehensive approach that encompasses all of the structural, technological and human-social factors. Developing Social capital in organization has a direct and significant impact on implementing knowledge management. Social capital can enhance an organization's ability to manage knowledge. It can help to facilitate the development of collective intellectual capital as well as enhance knowledge capture, knowledge codification, and knowledge sharing. Social capital makes collective action more efficient.

This paper shows how social capital affects the development of an e portfolio application. The lack of taking into consideration of the social capital dimensions resulted in the ePortfolio failing to achieve its intended goals. The central premise of social capital is that social networks have value. Social capital refers to the collective value of all "social networks" [who people know] and the inclinations that arise from these networks to do things for each other ["norms of reciprocity"]. The important reason is that knowledge management is a highly human social process and the nature of interactions among people has a great impact on the creating and sharing knowledge.

Finally, the authors believe that the SCKM approach may contribute as an alternative mechanism in designing systems which involve co-creation of value between the designer-developers and the users-owners of the system for both win-win output. Future works can explore the use of SCKM as analytical tool for social networked applications.

Acknowledgments. This study is funded by Advanced Informatics School, Universiti Teknologi Malaysia (UTM AIS) and Flagship Grants PY/2014/02737, PY/2014/02742, PY/2014/01542, PY/2013/00844 and PY/2014/02786.

References

Adler, P.S., Kwon, S.-W.: Social capital: prospects for a new concept. Acad. Manag. Rev. **27**(1), 17–40 (2002)

Banks, B.: e-Portfolios: their use and benefits. A white paper. Version 1. (2004). http://eel.sandbox.core-ed.org/content/download/382/1832/file/eportfolios_their_use_and_benefits.pdf

Barrett, H.: White paper: researching electronic portfolios and learner engagement. Accessed 23 June 2006 (2005)

Belaid Kridan, A., Steven Goulding, J.: A case study on knowledge management implementation in the banking sector. Vine **36**(2), 211–222 (2006)

Blanchard, A., Horan, T.: Virtual communities and social capital. Soc. Sci. Comput. Rev. **16**(3), 293–307 (1998)

Bourdieu, P., Wacquant, L.J.: An Invitation to Reflexive Sociology. University of Chicago Press, Chicago (1992)

Burt, R.S.: Structural Holes: the Social Structure of Competition. Harvard University Press, Cambridge (1992). http://books.google.com/books?hl=en&lr=&id=FAhiz9FWDzMC&oi=fnd&pg=PR5&dq=Structural+Holes:+The+Social+Structure+of+Competition&ots=vIL4hz4OLN&sig=UM03tgElUuMf2zsnrJka3eViafo

Chang, C.-C., Tseng, K.-H., Yueh, H.-P., Lin, W.-C.: Consideration factors and adoption of type, tabulation and framework for creating e-portfolios. Comput. Educ. **56**(2), 452–465 (2011)

Chen, H.L., Black, T.C.: Using e-portfolios to support an undergraduate learning career: an experiment with academic advising. Educause Q. **33**(4) (2010)

Child, J., Rodrigues, S.: The role of social identity in the international transfer of knowledge through joint ventures. Politics Manag. Knowl. **46**(68), 133–139 (1996)

Chinying Lang, J.: Social context and social capital as enablers of knowledge integration. J. Knowl. Manag. **8**(3), 89–105 (2004)

Coleman, J.S.: Social capital in the creation of human capital. Am. J. Sociol. S95–S120 (1988)

Davenport, T.H., Guest, V.G.: Special issue: knowledge management. J. Manag. Inform. Syst. **18** (1), 3–4 (2001)

Drucker, P.F.: Post-Capitalist Society. Routledge (1994). http://books.google.com/books?hl=en&lr=&id=IYVBmM5z69cC&oi=fnd&pg=PP5&dq=peter+drucker+1993+post-capitalist+society&ots=wF7HWiXA_Z&sig=PQB-XNOwFVnA48QyyUil_8UJr3I

Edvinsson, L.: Developing intellectual capital at Skandia. Long Range Plan. **30**(3), 320–373 (1997)

Eynon, B., Gambino, L.M., Török, J.: What difference can ePortfolio make? A field report from the connect to learning project. Int. J. ePortfolio **4**(1), 95–114 (2014)

Fukuyama, F.: Trust: The Social Virtues and the Creation of Prosperity. Free Press, New York (1995)

Hoffman, J.J., Hoelscher, M.L., Sherif, K.: Social capital, knowledge management, and sustained superior performance. J. Knowl. Manag. **9**(3), 93–100 (2005)

Jaview Carrillo, F., McElroy, M.W., Jorna, R.J., van Engelen, J.: Rethinking social capital theory: a knowledge management perspective. J. Knowl. Manag. **10**(5), 124–136 (2006)

JISC: “Effective Practice with e-Portfolios” publication (2008). http://www.jisc.ac.uk/media/documents/publications/effectivepracticeeportfolios.pdf. Accessed 27 March, 2013

Kogut, B., Zander, U.: Knowledge of the firm and the evolutionary theory of the multinational corporation. J. Int. Bus. Stud. 625–645 (1993)

Kogut, B., Zander, U.: What firms do? Coordination, identity, and learning. Organ. Sci. **7**(5), 502–518 (1996)

Krebs, V.: Social capital: the key to success for the 21st century organization. IHIRM J. **12**(5), 38–42 (2008)

Lesser, E.L.: Leveraging social capital in organizations. Knowl. Soc. Capital: Found. Appl. **3**, 16 (2000)

Lewick, R., Bunker, B.B.: Developing and maintaining trust in work relationships. Trust in Organizations: Frontiers of Theory and Reach, pp. 114–139 (1996)

Manning, P.: Explaining and developing social capital for knowledge management purposes. J. Knowl. Manag. **14**(1), 83–99 (2010)

Merton, R.K.: Social theory and social structure. Simon and Schuster (1968). http://books.google.com/books?hl=en&lr=&id=dyqZOcux9o0C&oi=fnd&pg=PR7&dq=Social+Theory+and+Social+Structure&ots=_PoL42pYmz&sig=7MaasjCkpIUVv86KRfZV70cb2u0

Monavvarian, A., Asgari, N., Akhavan, P., Ashena, M.: Developing social capital for facilitating knowledge management practices. Int. J. Soc. Econ. **40**(9), 826–844 (2013)

Nahapiet, J., Ghoshal, S.: Social capital, intellectual capital, and the organizational advantage. Acad. Manag. Rev. **23**(2), 242–266 (1998)

Pearlson, K., Saunders, C.S.: Managing and Using Information Systems: A Strategic Approach. Wiley, New York (2004). http://www3.cis.gsu.edu/dtruex/courses/MBA8125/Lectures-pdf/Day9-ProjectManagement-PearlsonCh11.pdf

Putnam, R. Social capital: measurement and consequences. Can. J. Policy Res. **2**(1), 41–51 (2001)

Putnam, R.D.: The prosperous community. Am. Prospect **4**(13), 35–42 (1993)

Ravet, S.: ePortfolio for a learning society. In: eLearning Conference, Brussels, pp. 19–20, May 2005. http://www.eife-l.org/activities/projects/epicc/final_report/WP7/EPICC7_7_Paper%20Brussells%20S%20Ravet.pdf

Sandefur, R.L., Laumann, E.O.: A paradigm for social capital. Rationality Soc. **10**(4), 481–501 (1998)

Smedlund, A.: The knowledge system of a firm: social capital for explicit, tacit and potential knowledge. J. Knowl. Manag. **12**(1), 63–77 (2008)

Sullivan, P.H.: Profiting from intellectual capital. J. Knowl. Manag. **3**(2), 132–143 (1999)

Teece, D.J.: Capturing value from knowledge assets. Calif. Manag. Rev. **40**(3), 55–79 (1998)

Tosh, D., Light, T.P., Fleming, K., Haywood, J.: Engagement with electronic portfolios: challenges from the student perspective. Can. J. Learn. and Technology/La Revue Canadienne de L'apprentissage et de La Technologie **31**(3). http://www.cjlt.ca/index.php/cjlt/article/viewArticle/97. (2005)

Turban, E., Leidner, D., McLean, E., Wetherbe, J.: Information Technology for Management, (With CD). Wiley, New York (2008)

Zainal Abidin, W.: E-Porfolio at UTM: a reflection on its journey (2008). http://eprints.utm.my/19842/

Zainal-Abidin, W., Uden, L., Alias, R.A.: Electronic portfolio as a knowledge management tool: a comparative analysis. In: The 8th International Conference on Knowledge Management in Organizations, pp. 179–188. Springer, Heidelberg (2014). http://link.springer.com/chapter/10.1007/978-94-007-7287-8_15

Management of Cooperation Activities in University Science Park

Anna Závodská[1(✉)], Zuzana Juchová Jurošková[1], and Jakub Soviar[2]

[1] University Science Park, University of Žilina, Univerzitná 1, 010 26 Žilina, Slovakia
{anna.zavodska, zuzana.juchova}@uvp.uniza.sk
[2] Faculty of Management Science and Informatics, University of Žilina, Univerzitná 1, 010 26 Žilina, Slovakia
jakub.soviar@fri.uniza.sk

Abstract. Cooperation is crucial for organizations in order to remain competitive. However, not all cooperative activities companies, universities or science parks do are successful. It always needs to be a win-win partnership for both parties. For cooperation to be successful, it is important to effectively use available resources and technologies. This paper investigates the cooperative activities in the University Science Park of the University of Žilina in Slovakia. The study shows that University Science Park failed in its cooperative activities due to several reasons. Therefore the purpose of this paper is to analyze the current cooperative activities of the University Science Park in Žilina as well as propose future activities towards improved cooperation in order to ensure its further successful operation. A matrix based on the cooperative principles was proposed.

Keywords: University science park · Project · University · Transport · Information system · Cooperation

1 Introduction to the Topic: Science Parks and Cooperative Management

Slovakia has no tradition in building and operating science parks at all comparing to any developed country where the science parks operate for decades. Because there are no science parks in Slovakia, technology transfer was done more intuitively by universities or individual researchers. Moreover, applied research is lagging behind and commercialization of research results has no successful outcomes.

As a results of the problems mentioned above there have been strong discussion for the past few years about building science parks which will increase innovation and applied research at the universities across Slovakia. After many years of discussions in 2013 six Slovak universities received funding from the European Union to build science park or research center. The University of Žilina got both of the projects approved. However, these projects will finish at the end of 2015. After this date the University Science Park has to be sustainable for the following 5 years.

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 770–779, 2015.
DOI: 10.1007/978-3-319-21009-4_56

We believe that cooperative activities with strong accent on building research networks of experts around the world may be a key to its success. However, without motivation of individual researchers to innovate and conduct applied research it will not be possible therefore we need to focus our attention on the human factor and create environment of strong research activity based on building international cooperation, excellent research networks and creating joint projects.

The main aim of this paper is to gain new knowledge in the field of science parks with a focus on identification of cooperative activities towards improving applied research as well as project cooperation.

Qualitative research was used as core approach for data collection. We used ***case study method*** for analyzing, summarizing and evaluating the current collaboration initiatives at the University Science Park and ***content analysis*** of external secondary data from websites concerning science parks and their collaboration activities (used for analyzing what does exist in different science parks across the world).

Research is rapidly progressing therefore collaboration with researchers from various fields as well as various universities is important. The purpose of this study was to define main cooperation areas and processes and set recommendation for its management. The data was collected over a year.

This paper begins with brief review of cooperative management and science parks. Subsequent section describes the case study of the University Science Park. This is followed by analysis of its cooperative activities and recommendations for their improvement. The paper concludes with suggestions for further research.

2 Cooperative Management and Science Parks

Science parks are environment where intensive knowledge transfer processes takes place. These processes are significantly connected with university research and companies to which the results of research are commercialized. Universities and companies have their own goals and science parks stand in the middle of these goals and manage all their main activities according to the main stakeholder's terms and situation. These processes are the main concern of science park's management – they are situated between independent companies and universities research bodies. They are not under a full control of a single organizational structure. They need to be managed in dynamic environment, on the market, in dynamic way. We stated this art of management as cooperative management – a wide theoretical and empirical background for our topic.

We define cooperative management as an "effective and efficient management of relationships in cooperation between separate and relatively independent organizations or individuals, with the goal of improving their competitiveness" [14]. This definition is based on our previous findings (e.g. [18–21]). Other authors in the field of cooperation management see this subject in a similar way. For example Lafleur [5] defines cooperative management as a way of managing and developing collaboration in a competitive environment. Ray [11] emphasizes that cooperative management represents a term for integrated management of company networks. Staatz [16] sees cooperative management as cooperative decision making within heterogeneous preferences. He highlights the need for a model of cooperation based on a defined group choice. Zhang [23] underlines that cooperative management represents a basis for solving all managerial problems. According to him, cooperative management provides conditions for creating a system of cooperation based on effective use of resources and technologies.

Clusters, science parks, science and technology parks or incubators can be a good example of cooperation with direct impact on increasing innovation [6]. Very important in this case is cooperation of universities with practice [7]. Thanks to that kind of cooperation transfer of technology can be achieved. Knowledge management can be a helpful tool in technology transfer process [8].

Science Parks. The generation of new ideas and their commercialization has traditionally been done internally, and companies rarely resorted to sharing innovative results, believing this could adversely affect their ability to generate competitive advantage [2].

On the other hand, if science parks do not share results of their research they might lose their competitive ability.

A science park is: "an organization managed by specialized professionals, whose main aim is to increase the wealth of its community by promoting the culture of innovation and the competitiveness of its associated businesses and knowledge-based institutions. To enable these goals to be met, a Science Park stimulates and manages the flow of knowledge and technology amongst universities, R&D institutions, companies and markets; it facilitates the creation and growth of innovation-based companies through incubation and spin-off processes; and provides other value-added services together with high quality space and facilities [3]."

Areas of innovation, of which science, technology and research parks are a highly specialized type, play a key role in the economic development of their environment. Through a dynamic and innovative mix of policies, programs, quality space and facilities and high value-added services, they [4]:

- "stimulate and manage the flow of knowledge and technology between universities and companies.
- facilitate the communication between companies, entrepreneurs and technicians.
- provide environments that enhance a culture of innovation, creativity and quality.
- focus on companies and research institutions as well as on people: the entrepreneurs and 'knowledge workers'.
- facilitate the creation of new businesses via incubation and spin-off mechanisms, and accelerate the growth of small and medium size companies.
- work in a global network that gathers many thousands of innovative companies and research institutions throughout the world, facilitating the internationalization of their resident companies."

The Association of Science and Technology Parks of Spain (APTE) defines Park as a Project, generally associated with a physical space, with the following characteristics [1]:

- "Formal and operational dealings with universities, research centers and higher educational institutions.

- Designed to encourage the creation and growth of knowledge – based companies and other organizations belonging to the service sector, which are normally established in the park itself, with a high added value.
- A stable managing body that promotes the transfer of technology and fosters innovation between the companies and organizations using the park."

A science park is a business support and technology transfer initiative that [17]:

- "encourages and supports the start-up and incubation of innovation-led, high-growth, knowledge-based businesses.
- provides an environment where larger and international businesses can develop specific and close interactions with a particular center of knowledge creation for their mutual benefit.
- has formal and operational links with centers of knowledge creation such as universities, higher education institutes and research organizations."

On the basis of the above definitions of science parks we analyzed our science park from the cooperative perspective.

3 University Science Park: Main Findings

This paper investigates the University Science Park which is an organizational unit of the University of Žilina in Žilina and moreover, it is an EU project until the end of 2015. Literature review and secondary research showed that cooperative activities in science park are very important. Nevertheless, the University Science Park cooperative activities are not well managed and therefore not creating any value, any innovation or any revenues. Furthermore, the University Science Park is not collaborating with external partners enough which is significantly decreasing its possibility to succeed on the market. Due to this fact University Science Park management should revaluate its attitude towards cooperation.

The role of the University Science Park is the realization of the applied research within the cooperation of the University Science Park researchers and experts from praxis on the national and international level. The focus of its activities is closely interconnected with the direction of the University of Žilina in Žilina whose direct part is the University Science Park.

The main activities of applied research of the University Science Park are divided into four divisions. Each division represents a separate functional unit with own organizational structure and management. Comprehensive supervision of their activities is carried by the director of the University Science Park as well as the rector of the University of Žilina. The divisions are following:

- Information and communication technologies (ICT),
- Intelligent manufacturing systems (IMS),
- Intelligent transport systems (ITS),
- Advanced materials and technologies (AMaT).

ICT division has a special position because in addition to solving its own research activities it also provides support for the use of ICT for conducting the activities of other divisions, especially IMS and ITS.

The largest division in terms of research tasks spectrum, research infrastructure and staffing is ITS. In the field of road traffic the University Science Park cooperates with its main project partner - Transport Research Institute, Inc.

An individual department of the University Science Park, which is working independently as other divisions, is the Centre for Technology Transfer (CTT). This Centre has its own organizational structure and the direct supervision of its activities is carried out by the director of the University Science Park. The CTT provides its services primarily to divisions of the University Science Park and subsequently to other workplaces of the University of Žilina. It also serves to external subjects. The CTT activities must be conducted in accordance with the overall concept of the University Science Park. Therefore, the activities of the CTT are closely linked to research activities tackled under divisions of the University Science Park.

International cooperation. Authorized personnel of the CTT have been working on building international relations from the beginning of the University Science Park project. These personnel are largely dependent on contacts acquired in the past on which capitalization are now seeking.

An important event for the acquisition of a network of contacts that might be beneficial for the future activities of the University Science Park are major international conferences where it is possible to present the University Science Park, its activities and progress. Moreover, it is crucial to look for advices as well as cooperation possibilities there. The big disadvantage is that it is time-consuming and expensive due to the fact that building a strong and beneficial relationship with foreign enterprises and institutions requires mainly personal contact, especially at the initial level.

University Science Park is currently mainly focused on foreign research institutions that have potential to collaborate on international projects. Another challenge is to look for the partners to cooperate on projects within the Horizon 2020 Framework Program.

At the same time we constantly work on the preparation of materials addressed to foreign enterprises to whom we can commercialize results of applied research of the University Science Park or we could make a deal to work on contract research.

Currently we have few partners that we intensively cooperate with, e.g. universities in Finland, Chile, Germany, Taiwan, and Belgium. We are working on new agreements with universities in Canada, and the USA.

When travelling abroad for a conference it is always important to contact universities involved in such projects as the University Science Park is so they can help with advices on how to build it. In this particular case the visit of Taiwan was very fruitful. Thanks to the contacting of our representative there we were able to visit science parks in Taiwan which are otherwise closed to public. From this initiative ministries of Slovakia and Taiwan started to closely cooperate and we are always informed about new steps they are taking. Other example is Chile where we have close partnership with one private university. They have many initiatives towards increasing applied research and entrepreneurship of students as well as researchers. Nowadays we have been deciding on which call from Horizon 2020 would be best for preparing new joint project with Chile and other partners.

Cooperation within the region. Žilina Region as well as the University of Žilina have been closely linked to the area of transport and ICT for decades.

There have been number of meetings with representatives of enterprises from various fields held where the new areas of common research interest have been searched in order to start new joint collaboration.

The meetings has indicated that the research focus of the University Science Park is in the right direction in terms of new trends in research so it makes sense to focus our effort on these particular research areas (represented by divisions).

The UNIZA received funding for ERA Chair in Intelligent Transport Systems. Excellent scientists could apply for the ERA Chair position established at the University Science Park of the University of Žilina within the prestigious European grant awarded to only 11 European research institutions within the Seventh Framework Program. The project aims to enhance research and innovation aspects of the park in the field of Intelligent Transport Systems (ITS).

The University of Žilina plays a leading role in the ITS research and development in Slovakia and it is equipped with state-of-the-art ITS and ICT infrastructure.

Cooperation with companies. In the Žilina Region there are operating many institutions assisting in development of ICT cluster, which was created in 2008. The ICT cluster Z@ict was established with the aim to improve quality of the life of Žilina Region inhabitants, increasing its prosperity and attractiveness as well as promoting the competitiveness of institutions and companies operating in the field of ICT and related services in the Žilina self-governing region. [22]

The University of Žilina has been actively cooperating with many companies from the ICT sector for ages, e.g. Scheidt & Bachmann, s. r. o., Siemens, s.r.o., IPESOFT, s.r.o., IBM Slovakia, s.r.o., AT&T Global Network Services Slovakia, EMTEST, a.s., Kros, a.s., etc. This cooperation is also beneficial for students to whom the companies are providing the possibility for internship or to work on their thesis (bachelor, diploma, and dissertation) on the selected topic that the particular company provided to students. These dissertations are focused on real practical problems that companies face. These companies also support students in form of providing sponsorship to various non-profit events (e.g. Startup Weekend, Rails Girls).

This type of cooperation the University of Žilina aims to support in the future and also strengthen through new technologies, equipment, and research activities at the University Science Park.

4 Discussion

Based on the previous analysis we proposed the following recommendations to University Science Park.

Building the whole new collaboration model. The model should reflect on actual cooperation relations of the University Science Park. It should also reflect on future tasks, park's vision and strategic goals. This model is a tool for park's management to plan and manage cooperative activities. Therefore the model should be dynamic by its nature. This dynamics will ensure the environment adaptation to new tasks or to market change, etc.

Model construction framework:

(a) Definition of main cooperation stakeholders and their prioritization. An effective tool for this task is the "stakeholder management matrix" (e.g. in [12]).
(b) Setting up relations between stakeholders that should be primary oriented on management activities.
(c) Setting up an organizational matrix – organization structure and its role in cooperative relations. These settings should ensure dynamics in possible organization structure changes, ad hoc connections with partners in limited time, time limited participation on single projects, etc.
(d) Standard management tasks: duties and responsibilities; control mechanism, etc.

For discussion, we prepared a project of main stakeholder's management matrix (identified by our analysis) (see Fig 1).

Boundary connection. Stakeholders are characterized by high uncertainty and not crucial importance. We put potential partners for cooperation there. Science park management must actively identify them and develop future relations with them. This part of the matrix is primarily focused on marketing activities.

Environment uncertainty	*High uncertainty*	**Boundary connection** – Potential partners for cooperation: companies, clusters, R&D facilities…	**Partnership** – International projects (Horizon 2020, etc.) partners – Foreign partner universities – Cooperating companies and clusters
	Low uncertainty	**Monitor** – Žilina self-government region – Regional and other Slovak universities – Selected conferences – State of the art in selected areas of scientific research	**Management of stakeholders** – University of Žilina – Slovak ministry of education
		Important but not crucial	*Crucial importance*
		Stakeholders importance	

Fig. 1. Stakeholders management matrix (based on the concept of [12])

Monitor. Stakeholders situated in this area are not crucial but it is necessary to monitor their operation on the market and/or in selected environment. Park's points of interest like scientific conferences and overall state of scientific research, etc. are also in the monitor part of the matrix. Management activities are oriented on gathering and processing of actual information considering selected stakeholders. Primary function is to have relevant information for decision making process.

Partnership. In this area are situated stakeholders which are in a close relationship with the University Science Park. There are common projects, common research activities, science park business partners, etc. All of these stakeholders are independent organizations so the uncertainty of environment is high. Management main activities in this area must be focused on: serious work on ongoing projects; satisfying of customer's needs; building of mutual trust based on previous positive experience, etc.

Management of stakeholders. Very important stakeholders and low level of environment uncertainty are the last area of our matrix. Here are situated stakeholders which are strongly connected with the University Science Park and have most significant effect on its operation. They could affect also complex decision making processes, and are able to modify them. They are able to be a part of science park's strategic planning, etc. Main role of management must be oriented on building of strong and effective relationships with these stakeholders.

For bringing the proposed stakeholders management matrix into life it is necessary to follow basic common management principles (e.g. in [12]) as well as basic principles of knowledge management (e.g. [9, 10, 13]) and also possibilities created by its application (e.g. [15]). ***Main objectives*** of proposed matrix should be focused on its implementation into the science park's cooperation activities management:

(a) Systematic analysis of stakeholders.
(b) Use of theoretical knowledge and real goals for setting the management criteria for single matrix positions.
(c) Practical management of single stakeholder's positions.
(d) Feedback based on experience with single stakeholders.
(e) New setting of the matrix according to gained feedback and/or management goals.

Application of the proposed matrix into the science park's processes should be taken on strategic level of its management. Science park's CEO should use this draft and deliver competences to single employees to work with the structure. The CEO should provide strategic point of view and also the specification of single goals. Science park employees should provide analytical work in order to keep the matrix up to date and execute single steps following the matrix.

Another crucial factor leading to success of management of cooperative activities is effective ***information system***. In this case we proposed various features to University Science Park internal information system (Share Point) such as: management of documents related to international cooperation (agreements, invitation letters, etc.), interconnection with project indicators, business trips organized in order to make new cooperation agreements, warnings about the end of cooperation agreement, database of contact persons at institutions we cooperate with, etc. We believe that this information system will be a helpful tool for managing various types of cooperative activities in the University Science Park.

5 Conclusion

Slovakia is one of the countries where commercialization of research results is very poor due to the lack of applied research, technology transfer, and innovative ideas of researchers leading to intellectual property protection as well as insufficient research infrastructure. There have been extensive debates in Slovakia during the last few years if all these problems should be solved by building a science park where researchers can find the best environment for their research. However, the science park by itself is not enough for solving a problem of low applied research activities. Without motivation of researchers to conduct serious research which would be then commercialized even the best condition would be negligible. This problem is complex and requires changing more conditions. One of the conditions is level of cooperative activities.

This paper, therefore, investigates the current cooperative activities in the University Science Park of the University of Žilina. This investigation shows that although the University of Žilina has a big project for building the University Science Park, cooperative activities are not yet sufficient enough. We found out that cooperative activities are crucial in building science park. It is because the research networks may be potentially growing by the good management of cooperative activities as well as more projects can be gained from the European Union which will lead to better conditions for performing research. One crucial thing is also having enough business partners with which researchers can work on applied research.

Based on our analysis, we proposed stakeholders management matrix. This matrix should help managers in leading their cooperative activities in right direction. We believe that this proposed matrix can be implemented into the University Science Park. In order to verify this matrix, further empirical studies will be needed.

Acknowledgements. This paper is supported by the following project: University Science Park of the University of Žilina (ITMS: 26220220184) supported by the Research&Development Operational Program funded by the European Regional Development Fund and by the Slovak scientific grant VEGA 1/0621/14 Marketing management in cooperative environment – Proposal of strategic cooperation management implementation model.

References

1. APTE. Definition of Park. (2014). http://apte.org/en/definition-of-park.cfm. (Accessed 24 September 2014)
2. Barrett, M., Velu, C.H., Kohli, R., Salge, T.O., Simoes-Brown, D.: Making the transition to collaborative innovation: Issues of readiness, trust and governance. Nesta Business Briefing, May 2011
3. IASP. Science park (IASP official definition) (2014 a). http://www.iasp.ws/knowledge-bites. (Accessed 24 September 2014)
4. IASP. The role of STPs and areas of innovation (2014 b). http://www.iasp.ws/the-role-of-stps-and-innovation-areas. (Accessed 24 September 2014)

5. Lafleur, M.: A Model for Cooperative Challenges. Cooperative Grocer Network, No. 116 (2009). http://www.cooperativegrocer.coop/articles/2009-01-21/model-cooperative-challenges. (Accessed 17 November 2014)
6. Mikušová, M.: Inovačné nástroje regionálneho rozvoja (Innovative tools for regional development). In: Rozvoj Euroregiónu Beskydy V. – úloha samospráv a inštitúcií pre rozvoji regiónov. Zborník z medzinárodnej vedeckej konferencie. Žilina, Bialsko – Biala, str. 124 – 132(2008), ISBN 978-80-8070-988-4 (in Slovak)
7. Mikušová, M.: Vzdelávanie založené na rozvoji kompetencií (Education based on development of competences). In: MARVI - Marketing vzdelávacích inštitúcií. Zborník z medzinárodnej vedeckej konferencie. Žilina, str. 112–117, November 2008, ISBN 978-80-8070-921-1 (in Slovak)
8. Mikušová, M.: Znalostný manažment ako integrálna súčasť podnikovej kultúry (Knowledge management as an integral part of company culture). In: Manažment ľudského potenciálu v podniku - zborník príspevkov 6. medzinárodnej vedeckej konferencie. Žilina, 19–20 máj 2009 Žilinská univerzita, str. 97–102, ISBN 978-80554-0013-6 (in Slovak)
9. Nonaka, I., Konno, N.: The concept of Ba: building a foundation for knowledge creation. Calif. Manag. Rev. **40**(3), 40–54 (1998)
10. Nonaka, I., Toyama, R., Hirata, T.: Managing Flows – a Process Theory of the Knowledge-Based Firms. Palgrave Macmilian, London (2008)
11. Ray, P.K.: Cooperative Management of Enterprise Networks. Kluwer Academic Publishers, Dordrecht (2002)
12. Robbins, P.S., Coulter, M.: Management. Grada, Praha (2004). ISBN 80-247-0495-1
13. Schwartz, G.D., et al.: Encyclopedia of Knowledge Management. Idea, London (2006)
14. Soviar, J.: From Cooperation to Management – Cooperative management. Habilitation thesis. University of Zilina, Faculty of Management Science and Informatics (2012) (in Slovak)
15. Soviar, J., Závodská, A.: Knowledge and its creation – the case of introducing product to the market. Verslas: Teorija ir praktika - Bus. Theo. Pract. **12**(4), 362–368 (2011). ISSN 1648-0627 print / ISSN 1822-4202
16. Staatz, J.M.: The cooperative as a coalition: a game-theoretic approach. Am. J. Agric. Econ. **65**(5), 1084–1089 (1983)
17. UKSPA. About UKSPA. (2014). http://www.ukspa.org.uk/our-organisation/about-us. Accessed 24 September 2014
18. Vodák, J., Soviar, J., Lendel, V.: Identification of the main problems in using cooperative management in Slovak enterprises and the proposal of convenient recommendations. Commun. – Sci. Lett. Univ. Žilina **15**(4), 63–67 (2013)
19. Vodák, J., Soviar, J., Lendel, V.: The evaluation system proposal of the businesses preparedness for cooperative management implementation. Verslas: Teorija ir praktika – Bus. Theo. Pract. **14**(4), 315–322 (2013). ISSN 1648-0627 print / ISSN 1822-4202
20. Vodák, J., Soviar, J., Lendel, V.: Cooperation management in Slovak enterprises. In: 2nd World Conference on Business, Economics and Management -WCBEM 2013. Procedia - Social and Behavioral Sciences, vol. 109, pp. 1147–1151, 8 January 2014, ISSN 1877-0428
21. Vodák, J., Soviar, J., Lendel, V.: Identification of the main aspects of cooperation management and the problems arising from their misunderstanding. Commun. – Sci. Lett. Univ. Žilina **16**(3), 93–101 (2014)
22. ZAICT. The aim of the association (2008). http://www.zaict.sk/. (Accessed 20 November 2014)
23. Zhang, W.: Cooperation system constructing and model of its operation mechanism. In: Proceedings of the International Conference on Business Management and Electronic Information (BMEI), vol. 3, pp. 784–787 (2011)

Management Model for Planning Articulation at Universidad Distrital Francisco José de Caldas

Jenny A. Triana[1(✉)], José I. Rodríguez[2], and José I. Palacios O[2]

[1] Posgrados, Universidad Escuela Colombiana de Carreras Industriales -ECCI, Bogotá, Colombia
jtrianac@ecci.edu.co

[2] Facultad de Ingeniería, Universidad Distrital Francisco José de Caldas, Bogotá, Colombia
{jirodriguez,jpalacios}@udistrital.edu.co

Abstract. The management system at Universidad Distrital Francisco José de Caldas has not allowed the achievement of the strategic objectives set in the Strategic Development Plan 2008-2016 "High Impact Knowledge, Expertise and Research for the Human and Social Development". The lack of articulation of the already defined strategy, along with the activities developed by the different units, require the reformulation of a system that allows the execution and assessment of the results in terms of the University's mission statement. The implementation of the Hoshin Kanri Management System, based on the Deming cycle (planning, do, verify, and act) shall guarantee that upon a long-term plan, the objectives of the University's top management are articulated with objectives at a lower hierarchical level through a cascade deployment process, so that actions merge with the strategy and the results will consolidate a high-impact University focused on research to solve the problems of the city-region of Bogotá and the country itself.

Keywords: Management model · Ten-year development plan · Planning model · Hoshin kanri strategic planning

1 Introduction

The Budget allotted to a public university is the main and most powerful individual controller of activities. However, higher education institutions (HEI) in Colombia, after having a period of certainty, adopted the techniques of strategic planning towards the development of their institutional horizon, by preparing development plans and break budget limitations. But not all the HEIs did so. It is the case of Universidad Distrital Francisco José de Caldas (UDFJC), which is concerned about facing the new challenges within the public education context, and therefore created the Strategic Development Plan 2008–2016: "Knowledge, Expertise and Research for the Human and Social Development", hereinafter referred as PED.

The lack of articulation of the already defined strategy, which is part of the ten-year development plan, along with the activities that have already been programmed and

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 780–793, 2015.
DOI: 10.1007/978-3-319-21009-4_57

executed by the several units of the university, has been the biggest obstacle in the way to assess the results of activities in terms of the mission statement of the university and the development of its institutional horizon. In other words, at the end of each term, the management plan accounts for the results achieved which are different from the strategic objectives.

The consequences affect the strategic alignment directly as it separates the units and processes in such manner that all the work, actions and decisions made at all levels do not take link the purposes and the objectives and shatters the vision statement of the University.

Big companies around the world have faced similar situations, and in order to control damage, the Hoshin Kanri Management System was created. Such system is based on the Deming cycle: plan-do-check-action. The system articulates the objectives of the top management with the objectives at lower positions, through a cascade deployment process so that actions merge in terms of strategy and results are as expected.

1.1 Background and Description of Issues

Bearing in mind the commitment of UDFJC with society and the impact expected in its scope of influence, meeting Article 36 of the General Statute that presents the CSU as the instance that establishes the "planning system whereby the institutional project is prepared, executed and assessed to the Academic Council for the realization of principles, achievement of objectives and development of the University management" [1]; and by exercising the university's autonomy, the PED (2007–2016) was prepared and executed through Agreement No. 001 of 2008 of the CSU. Its approach defines the institutional priorities for the growing and development, allowing "the guidance of management and the coordinated action of the several actors bound to the development of the University in the long term" [2], based on the idea defined by the UDFJC (2007) in its PED: "Knowledge, Expertise and Research for the Human and Social Development", the University started to give priority to a series of projects in its three-year indicator plan.

The articulation of the strategy defined, if taken into consideration on a daily basis, it is inexistent. It is to say that there is not alignment between the activities developed by the administrative units of the university, the objectives of the processes and the future requirements. This is proved in the Progress Report of the management indicator plan (CBN 1013) which accounts for the level of compliance of PED up to December 2012. Because of the above, the difficulty to execute activities and assess results, both in terms of the mission statement and in terms of support, and therefore regarding timely actions established to guide them, arises. As a consequence, the results show that during the past years, the progress of UDFJC in terms of strategic objectives are minimum in relation to the initial projection. The above mentioned, although the Entity prepared an indicator plan (three-year plan) to facilitate the execution of projects, define possible arising emergencies, provide strategic monitoring and control that would allow closing the gap between strategic planning and plan operability, did not work.

Based on this study, we look for the Hoshin Kanri (HK) methodology to be a theoretical point of reference as it proposes a management model for planning articulation at UDFJC. However, the scope of the study will be the design of a proposal for a Planning Model; its implementation and results will depend on the adoption decision on the part of the institution.

2 Hoshin Kanri Origin and Guiding Principles

Back in 1954, Joseph Juran along with Edwards Deming worked on the concept of Statistical Quality Control (SQC) that started to be used in Japan by the end of the 1940s [3], and then in the 1970 s it became in what we nowadays know as Total Quality Control (TQC). By that time, the concept of Management by Objectives (MBO) broke in Japan and such ideas were combined with Deming and Juran's and then creating what we call "quality strategic planning" [4].

It was during that decade in which the Japanese continued improving their planning activities and in 1968, the Bridgestone Tire Company applied the Deming Cycle or PDVA (planning-doing-verifying-acting) with the participation of all the employees of the organization, established annual priority objectives determined in agreement with the annual policies of the company and related to the inter-functional management that time later were implemented in the relevant units of the organization. The top management diagnosed the activities in order to verify how things were being done, examined the results and identified any issue in the fulfillment of the objectives established. In such manner, the top management accompanied the policy improvement process. Then, the company named it as the "Hoshin Kanri" (HK) methoid. The HK methodology was a success in Japan by 1975 and was adopted nationwide [5, 6].

By mid 80s, the methodology reached the western world through Japanese subsidiary companies such as Hewlett– Packard and Fuji–Xerox. In the USA, the Hoshin Kanri methodology was successfully used by Florida Power and Light. In Europe, companies started to use it from the 90 s on as method to link their medium term strategies to their annual plans and achieved a remarkable performance improvement [7].

Currently, the HK methodology is used by most of the Japanese companies that operation globally and by western companies that use it under a different name, with some variations, but keeping the core idea as the center of the principles and method [8]. For instance, Bank of America –Hoshin Planning, Procter & Gamble –Policy Management, Xerox Corporation- Managing for Results, Unilever –Strategy into Action, Hewlett Packard –Hoshin Planning, AT&T –Policy Management, Donnelly –Managing for Results, and Exxon Chemical –Policy Management.

The HK methodology was developed as a TQC tool to achieve company flexibility and diminish response timing to changes in the business environment, becoming as one of the fundamental pillars for success in terms of TQC [9–15]. In this matter, the HK methodology is introduced to fight adversity and failures within the organization such as de-linkage of objectives established by the top management and the daily objectives at operational levels [16] or that even because of the sophisticated planning processes used, plans are different to what it is actually achieved.

On its part, the HK methodology allows the strategy management based on the several levels of the company and the several hierarchical duties, making it possible the linkage of the efforts of the organization as a whole to achieve the key objectives for the business [15]. The main principle on which the HK methodology is supported is that each one of the employees of the organization, no matter their duties or level must be contribute to fulfill the key objectives of the company for the success of the whole organization [17]. According to Kondo [18], the key points of the HK methodology are: 1. The establishment of annual guidelines; 2. The establishment of quality guidelines; 3. The transformation of such methodological guidelines in objective guidelines set by goals, objectives and strategic priorities, and 4. The top-down and bottom-up deployment of strategies and objectives to the entire organization by applying the PDCA (Plan-Do-Check-Action), to align all the levels [19], as its implementation in the policy and objectives established makes the success probability higher as every single level in the organization is improving its performance. This is the ideal methodology to efficiently achieve objectives by creating strategies to solve problems and action plans (PLAN), implementing such plans on a constant basis (DO) and checking if the results are as planned through assessment (CHECK). If expected results are not achieved, then measures to avoid mistakes is established. If results are as expected, then processes are standardized (ACTION).

This management cycle constitutes the HK Methodology basis that arises as a managerial philosophy that, through a participatory process, intends for establishing, deploying and auto-controlling the fundamental goals of the organization and its top management, at the same time that guarantees the necessary means and the resources to ensure that such goals are achieved at all levels of the organization [20].

3 Comparative Study of the Several Planning Systems of Higher Education Institutions Regarding Academic, Administrative and Financial Aspects

In order to design a proposal of an articulation model for planning and management for UDFJC, based on the HK methodology, we consider the comparison between renowned higher education institutions nationwide to unify criteria in order to detect and select the trends and planning points of reference. The criteria included allow for the identification of those meaningful trends in administrative, financial/budgetary, academic and research contexts similar to the ones at UDFJC, mainly:

1. Official Colombian Universities (state universities).
2. Presence within the first twenty (20) places in the performance rankings of universities in terms of their mission statement duties (research, educators, training and extended education and/or social projection), through:

   - *Sapiens research:* The study analyzes and measures Colombian universities according research indicators, in terms of three variables: indexed magazines in Publindex, postgraduate programs and research investigation teams in those universities. Results are shown in Table 1.

Table 1. Ranking U-Sapiens 2013

Rk	Higher education institutions (HEI)	Type	Department
1	Universidad Nacional de Colombia	Official	Bogotá
2	Universidad de Antioquia	Private	Antioquia
3	Universidad de los Andes	Private	Bogotá
4	Pontificia Universidad Javeriana	Private	Bogotá
5	Universidad del Valle	Official	Valle
6	Universidad Nacional de Colombia	Official	Antioquia
7	Fundación Universidad del Norte	Private	Atlántico
8	Universidad Industrial de Santander	Official	Santander
9	Universidad Pontificia Bolivariana	Private	Antioquia
10	Universidad de Caldas	Official	Caldas
11	Universidad Pedagógica y Tecnológica de Colombia	Official	Boyacá
12	Universidad Tecnológica de Pereira	Official	Risaralda
13	Universidad EAFIT	Private	Antioquia
14	Universidad Externado de Colombia	Private	Bogotá
15	Universidad de Cartagena	Official	Bolívar
16	Universidad Distrital Francisco José de Caldas	Official	Bogotá
17	Universidad de Medellín	Private	Antioquia
18	Universidad Santo Tomás	Private	Bogotá
19	Universidad del Cauca	Official	Cauca
20	Universidad de la Sabana	Private	Cundinamarca

Source. Ranking U- Sapiens (2013).

- *Integral Ranking of Universities BOT*: weighing in the ranking position of training quality (40 %), social extension (15 %), productive extension (15 %) and research (30 %) for HEI classified as universities. Results shown in Table 2.

3. Annual Budget higher than COP$120,000,000,000.
4. Broader coverage with more active students.

Based on the criteria aforementioned and having into account the results shown in Table 3, the following universities were chosen for the purpose of this proposal:

1. Universidad Distrital Francisco José de Caldas –UDFJC- (Source HEI)
2. Universidad Nacional de Colombia –UNAL-
3. Universidad de Antioquia –UDEA-
4. Universidad del Valle

3.1 Comparative Analysis

HEI responses to new demands in the knowledge society, understood as a society with a scientific approach, academized and focused on services, as mentioned by Bell [21] with a professional structure marked by the professionalized knowledge workers and

Table 2. Integral BOT university ranking -2013

Rk	Nombre IES	Quality training	Social extension	Productive extension	Investigation
		percentage weighting for integral ranking			
		40 %	15 %	15 %	30 %
1	Universidad de los Andes	1	3	4	1
2	Universidad Nacional de Colombia	5	1	2	2
3	Universidad CES	10	2	1	4
4	Colegio Mayor de Nuestra Señora del Rosario	2	6	18	6
5	Universidad EAFIT	6	5	3	14
6	Pontificia Universidad Javeriana	8	4	11	8
7	Universidad de la Sabana	7	10	5	11
8	Fundación Universidad del Norte	9	11	7	9
9	Universidad de Antioquia	13	22	6	3
10	Universidad ICESI	4	13	15	19
11	Universidad Industrial de Santander	12	18	40	7
12	Universidad del Valle	15	39	40	5
13	Universidad Externado de Colombia	3	26	22	38
14	Universidad Distrital Francisco José de Caldas	11	6	40	29
15	Universidad Pontifica Bolivariana	21	36	13	15
16	Universidad del Cauca	27	20	26	10
17	Universidad Tecnológica de Pereira	23	44	14	16
18	Universidad Militar Nueva Granada	18	51	10	26
19	Universidad Autónoma de Bucaramanga	25	14	24	30
20	Universidad de Caldas	32	21	40	13

Source. Ranking BOT (2013).

with academic qualifications, are having impacts on their structures and management [22, 23]. It is evident that these universities have designed transformation processes that look for dynamism and therefore their government and management processes have been changing for the good. That is why the reconfiguration of universities is not an easy task and requires the use of strategic planning to identify the main addressing of universities in such a way that the concentration of resources is provided by a limited number of duties in order to maximize the benefits for the external stockholders of the

Table 3. Selection criteria for universities to identify trends

University	N. Students (SNIES)	Annual budget (2013)	Sapiens (2013)	BOT (2013)
Universidad Nacional de Colombia	47.1	$ 1.040.722.735.327 (Resolución 1692)	1 y 6	2
Universidad de Antioquia	32.384	$ 312.915.000.000 (Acuerdo 402)	2	9
Universidad del Valle	26.01	$ 327.216.501.608 (Acuerdo 014)	5	12
Universidad Industrial de Santander	18.882	$ 298.777.176.100 (Acuerdo 078)	8	11
Universidad de Caldas	10.467	$ 164.882.490 (Acuerdo 056)	10	20
Universidad Pedagogica y tecnologica de Colombia	22.547	$ 129.287158.229 (Resolución 1313)	11	36
Universidad Tecnologica de Pereira	14.255	$ 124.152.000.000 (Acuerdo 36)	12	17
Universidad de Cartagena	13.047	$ 157.839.784.565 (Acuerdo 09)	15	38
Universidad Distrital Francisco José de Caldas	27.243	$ 253.483.255.000 (Resolución 046)	16	14
Universidad de Cauca	11.908	$ 149.630.064.110 (Acuerdo 074)	19	16
Universidad Militar Nueva Granada	9.826	$ 163.615.000.000	24	18

university, such as students, graduate students employers, financing organisms and society in general, as well as internal stockholders such as educators and administrative staff [15]. It is in this regard that the analysis of the strategic plans allows us to know how universities see themselves in the medium term and look for guaranteeing and improving efficiency and ensure their proper functioning.

For the purposes of this study, once the information related to the planning area has been systematized, the type of plan designed, its lifespan, its strategic axis, the management reports, the levels of budgetary execution, the management reports, the level of budgetary execution and the impact on the development of the strategy established for each one of the public universities and the plans developed, we can observe the use of best planning practices, as actions are systematic and have originated the identification of needs in HEI, in order to be more efficient and competitive.

From the strategic level, we have identified the following items as their common denominator: research, internationalization, financial and administrative modernization, university welfare and academic management. Their objectives were designed, after assessment, according to the particularities of each institution, their context and resources. However, most of them do not establish indicative and/or action plans which makes it difficult to organize daily endeavors in articulation with strategies.

Management reports analyze the impact of their development against strategy in a weak manner. Management reports limit their endeavors in relation to the mission statement of universities but are limited to record the documentary output, which in itself does not contribute to the development of plans. On the other hand, none of the institutions provided evidence on the formulation of their strategic plans based on the national and territorial development plans, although some of them are really strong locally or are biased in their region, their plans are not derived directly from this articulation, but from their own assessment within their context. In such regard, HEI subject matter of this study are in contraposition to Graffikin and Perry [24], who state that strategic plans differ according to the social circumstances of the university. In that way, some universities are really strong locally or are biased in their region that impact their activities and specialties, while others have broaden their activities to become international players [25–27].

For the purposes of this analysis, the plans designed may be interpreted as the main mechanism by which the influence of this organizational field are exercised, as it reflects the way the university community reacts to possible pressure from the environment, especially the legitimation pursued to obtain support from government and other organizations they depend on to obtain resources.

Although strategic planning processes are designed to be adapted to the specific needs of the corresponding University [28], there are some common features that were also analyzed, such as the idea of the university to first identify its vision and mission. This step was followed for a series of analysis of external and internal analysis and the gap between them, and was used to establish a point of reference in order to provide a context for strategic planning, including the development of strategies, goals and action plans. In addition, most of the universities submit management reports with the purpose of accounting for their activities to society, and show evidence of their transparency in the management of public resources. However, in Colombia is still common to find hierarchical organizational structures with several levels between the top management, the educators and students, which can actually have an impact on corruption practices as there may be formulation of strategies limited to particular interests or the non-compliance with institutional objectives.

4 Proposal of a Planning System for Universidad Distrital

The results of the comparative work identify a planning trend addressed to the mission statement using action plans, which were designed without being articulated with development plans or strategic plans, diminishing the possibilities to get HEI closer to their future-oriented plans and react quickly to changes.

In that manner, universities are organization that must be permanently adapted to changes and modernization processes for education, in order to keep being high quality institutions and competence [29].

Therefore, having into account that the main elements of the strategic planning and its application to HEIs, the opportunity to achieve integration items of key elements in strategic plans through a proposal of a planning system for the UDFJC based on the HK methodology has arisen. After this, a methodology for HK implementation is proposed using Catchball tools and the efficient meetings system.

The Hoshin planning system proposes a series of stages based on the PDCA cycle [30]. Starts with the analysis of the external and internal environment; then, the top management and/or managing director design a set of objectives, strategies and monitoring indicators.

After that, as mentioned by Liévano [31]:

> The persons in the next level of management, according to the objectives that have been established, establish a specific action plan for their area and so on until the lowest level set their objectives as well. Then, a project to cover the objectives flowing from the top to the bottom and their implementation teams is also established (p 33).

The following are the stages of the HK methodology processes:

1. Internal and external context analysis.
2. Establishment of desired future and objectives.
3. Preparation of an objectives deployment plan.
4. Plan implementation and monitoring.
5. Objective and plan assessment.

The HK methodology was developed as a TQC tool to achieve company flexibility and diminish response timing to changes in the business environment, becoming as one of the fundamental pillars for success in terms of TQC [11, 14, 15, 30, 32–34]. As during the formulation stage, the planning system of the Institution is robust and the results are coherent with the elements, both the planning and the strategy stages and the assessment process and its corresponding elements, the strategic planning system proposal for UDFJC is established as shown Fig. 1:

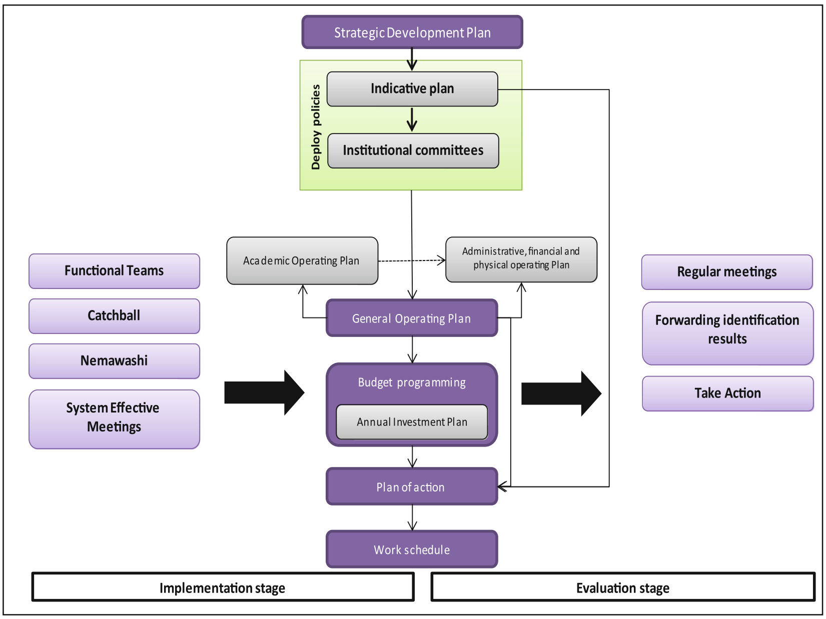


Fig. 1. Planning system proposal (source. authors)

4.1 Implementation Stage

The first step to implement the indicative or three-year plan aligned with the PED is to submit the guidelines for the three following years and deploy each of those guidelines among the 16 committees involved (Deans, Curricula, Admissions, Extended Studies, Research, Library, Laboratories, Institutional Self-Assessment, Publications, Welfare, Occupational Health, Inter-institutional Relations, Communications, Integrated Management Systems, and Facilities, equipment and telecommunications).

After that, having an intervention planned process, the need to introduce motivational elements addressed to officials (administrative staff and educators) arises, as well as introducing inter-functional teams to the organizational culture of the University. This must be done in order to ensure the development of the remaining elements such as Catchball tools, as defined by Witcher [35], a process of negotiation involving formal and informal meetings, translating the objectives in means, establishing the use of resources and converting objectives at different levels of an organization in annual goals agreed by all the members of the organization (p.218), the Nemawashi that allows the building of agreements even before involving the formal administrative processes, which avoids confrontations, promotes information interchange and team working [36], and the system of efficient meetings that makes emphasis on team working, under its motto: “for ever”, and its scientific approach to solve problems based on data and objective facts [30].

By adopting catchball tools, it is possible to introduce three-year and annual guidelines, deploy them in the different objective participating dependencies of the objective, which means the system requires deployment each three (3) years to articulate the governance program of the provost that has been elected and an annual deployment to establish an action plan including the objectives, goals, strategies, resources, people in charge, and measure parameters of processes upon adjustments made during the first deployment.

4.2 Evaluation Stage

After that, the result of the whole process is used as input for the following stage which is the permanently monitoring carried out through regular meetings: changes in the environment, deviations in the results of the indicators so that the teams can later identify the relevant resources in case of improvement points. Then, it is necessary to process and assign resources with the top management participating in the activity and led by the Provost, the Academic Council and University Superior Council to ensure the implementation of corrective and preventive actions. Finally, each year, an evaluation of the objectives and the plan must be conducted.

5 Conclusions and Future Papers

Public universities are immersed in transformation processes that force them to restructure their government and management systems in order to be dynamic and competitive. In such regard, the formulation and development of strategic plans has been spread to public universities to face global changes and challenges in terms of education.

Notwithstanding the above, in Colombia, Law 30 of 1992 has limited strategic planning to precarious budgets annually allotted to public universities, transforming HEI into tools to manage resources and the results are an answer to a mimetic and regulatory mechanism that only promotes the reinforcement of mission statements in some universities, and therefore the budget must become in the basis of public management rather than the object of the strategic plan, meaning that it is necessary to substitute the level of budgetary execution as the main indicator of excellent management.

On the other hand, it is necessary to have a planning system that guarantees the involvement of all the units of an institution to work on common objectives. Notwithstanding the above, public universities continue having the traditional hierarchical organizations with several levels between the top management, the educators and the students, which impacts the results as the decisions are made only by the top management without consulting or minding stakeholders.

The alignment between several stages will allow a higher level of efficiency in terms of the compliance and control of Strategic Development Plan [37], requiring the involvement of activities to establish policies and three-year objectives, annual objectives, allocation of resources and fit that information in an Operational Plan that would be effective if the deployment and communication of objectives is made by involving the university community in the process so that actions can be taken.

The foregoing would be possible by implementing the HK methodology, as it can improve the understanding of the requirements of users on the one hand, and on the other hand, periodic reviews contribute to a better control in the achievement of goals and higher flexibility to adapt themselves to the external influences without endangering the strategic plan and increasing the level of trust and motivation of human talent involved in the process.

As an answer to the research problem, we conclude that the planning system defined will provide elements that facilitate and articulate the strategy and the management, as well as the necessary adjustments resulting from new actions demanding the integration of methodologies through the HK methodology in the current planning system of the UDFJC.

The alignment between the different stages will allow a higher efficiency to comply with the PED. However, the implementation of the HK methodology at UDFJC may have a fundamental obstacle: the organizational culture. It is so because of apathy and fear to change.

If the PED of UDFJC was thought based on strategic objectives that have had serious difficulties to be materialized, the HK methodology will allow the University's strategy to revive thanks to its successive stages of explanation and commitment, always mediated by the use of catchball tools, making it possible the integration of strategic planning processes at high level involving missional and support processes. Such integration may be given because the HK methodology supports a solid methodology that guarantees the dialogue between the processes at different levels and supports the culture of agreement, team working and continuous improvement.

UDFJC may be a beneficiary of the HK methodology application as this methodology will harmonize wide strategies with high priority implementation plans, will guide the institution towards a selective set of priorities, will strengthen the use of data and information to identify strategic gaps for the focused improvement of activities, will foster the use of operational inter-disciplinary equipment to identify opportunities for improvement, will assign priorities and will promote the learning of the system.

For the successful implementation of the HK methodology, it is necessary to have certain administrative maturity which at this point the University does not have. Such maturity means the organizational culture that makes it difficult the adaptability to change, reinforcing the clear and constant presence of the top management in the implementation and development process of the HK methodology (key leadership for any change process), as currently the top management is only and poorly involved in the PED formulation and in an regulatory manner in the reporting of fulfillment indicators.

In terms of research, the points of reference analysis requirements were satisfied both for the strategic planning and the Hoshin Kanri methodology, as well as a design for a proposal of the strategic planning system of UDFJC according to the stages of the PDCA plan.

The following are the final conclusions:

1. The current planning system of UDFJC, despite the presence of strategic planning elements in the deployment process of policies and objectives along with the formulation of operational plans, does not have the tools that allow the articulation between the strategic planning and the management.
2. The organizational structure and the roles may fit. However, inter-functional teams and the efficient meetings system may facilitate the deployment of objectives as the corresponding monitoring mechanism.
3. In addition, the appropriation of Catchball or Nemawashi tools will contribute to improve the management and therefore the achievement of objectives established in the development plan.

The tool proposed will take UFJC to achieve a proper management on its strategic plan, complying with the provisions in the plan and safeguarding the resources allocated by the government and integrating the strategy, the academic future-plans, the operational activities and budgetary elements for investment and functioning that will allow the optimization of the resources and to have a functional structure that generates flexibility and rapidness for the university to adapt itself to the changing environment.

References

1. Universidad Distrital Francisco José de Caldas. Acuerdo 03 de Abril 8 de 1997. Por el cual se expide el Estatuto General de la Universidad Francisco José de Caldas (1997)
2. Universidad Distrital Francisco José de Caldas. Acuerdo 01 de Enero 18 de 2008. cual se confirma el Plan Estratégico de Desarrollo de la Universidad Distrital Francisco José de Caldas 2008–2016 Saberes, Conocimientos e Investigación de Alto Impacto para el Desarrollo Humano y Social, p. 1 (2008)

3. Lee, R., Dale, B.: Managing Quality. Prentice Hall, Londrés (1998)
4. Winiarski, P.: An introduction to Hoshin Kanri (2008). http://www.winenterprisesllc.com/How-To-Deploy-Strategic-Goals.pdf, Accedido el 20 de Enero de 2015
5. Kondo, Y.: Hoshin kanri: a participative way of quality management in Japan. TQM Mag. **10**(6), 425–431 (1998)
6. Tsung, M.Y., Chao, T.S.: Total quality in human resources management: an emerging perspective. Manage. Lab. Stud. **24**(2), 113–118 (1997)
7. Tennant, C., Roberts, P.: Hoshin kanri: implementing the cathball process. Longe Range Plann. **34**(3), 287–308 (2001)
8. Witcher, B.J., Butterworth, R.: Hoshin kanri: Policy management in Japanese-owned UK subsidiaries. J. Manage. Stud. **38**(5), 651–674 (2008)
9. Campos, V.F.: Controle da Qualidade Total. Qualitymark, Belo Horizonte (1996)
10. Akao, Y.: Hoshin Kanri: Policy Deployment for Successful TQM. Taylor & Francis, Boca Raton (2004)
11. Ibid, Lee y Dale
12. Jackson, T.L.: Hoshin Kanri for the Lean Enterprise: Developing Competitive Capabilities and Managing Profit. Productivity Press, New York (2006)
13. Dennis, P.: Fazendo acontecer à coisa certa: um guia de planejamento e execução para líderes. São Paulo. Lean Institute Brasil (2007)
14. Ibid, Winiarski
15. Ibid. Witcher
16. Ayala, N.F.: A Utilização Do Hoshin Kanri Para O Desdobramento Da Estratégica No Contexto Da Produção Enxuta(Tesis de Maestría), Universidad Federal de Rio Grande del Sur, Brasil (2010)
17. Witcher, B.J., Chau, V.F.: Balanced scorecard and hoshin kanri: dynamic capabilities. Manage. Decis. **45**(3), 518–538 (2007)
18. Ibid, Kondo
19. Yacuzzi, E.: La gestión Hoshin: modelos, aplicaciones, características distintivas. Serie de documentos de trabajo Universidad del CEMA. Buenos Aires, Argentina (2006)
20. Matamala, S.R.: Administración por políticas: Hoshin-Kanri –Bogotá. McGraw Hill, Colombia (1997)
21. Bell, D.: El advenimiento de la sociedad post-industrial. Un intento de pronosis social. Madrid, España: Alianza (2001)
22. Gumport, P., Pusser, B.: A case of bureaucratic accretion: context and consequences. J. High. Edu. **66**(5), 493–520 (1997)
23. Slaughter, S., Leslie, L.: Academic Capitalism: Politics, Policies, and the Entrepreneurial University. Johns Hopkins University Press, Baltimore (1997)
24. Graffikin, F., Perry, D.C.: Discourses and strategic visions: the U.S. research university as an institutional manifestation of neoliberalism in a global era. Am. Edu. Res. J. **46**(1), 115–144 (2009)
25. Clark, B.R.: The entrepreneurial university demand and response. Tert. Edu. Manage. **4**(1), 5–15 (1998)
26. Askling, B., Bauer, M., Marton, S.: Swedish universities towards self-regulation: a new look at institutional autonomy. Tert. Educ. Manage. **5**(2), 175–195 (1999)
27. Bauer, M., Askling, B., Gerard-Marton, S., Marton, F.: Transforming Universities: Changing Patterns of Governance, Structure and Learning in Swedish Higher Education. J. Kingsley, London (2000)
28. McConkey, D.: Strategic planning in nonprofit organizations. Bus. Q. **46**(2), 24–33 (1981)
29. Arias Montoya, L.: Cuadro de mando integral en Instituciones de Educación Superior Scientia et Technica. Universidad Tecnológica de Pereira **11**(27), 2 (2005)

30. Ibid, Akao
31. Liéavano, F.: Metodología de Integración de la Inteligencia Competitiva con el Sistema de Planeación Hoshin Kanri (Tesis de Maestría). Instituto Tecnológico y de Estudios Superiores de Monterrey, México p. 33 (2008)
32. Ibid, Campos
33. Ibid, Jackson
34. Ibid, Dennis
35. Ibid, Witcher, p. 218
36. Ibid, Yacuzzi
37. Universidad Distrital Francisco José de Caldas Informe de avance sobre el plan indicativo de gestión (CBN 1013) (2013)

Knowledge Ambassadors: Enhancing Tacit Knowledge Transfer in Kenyan Universities

Lucy Kibe and Tom Kwanya(✉)

Department of Information and Knowledge Management,
The Technical University of Kenya, Nairobi, Kenya
{kibelucy871, tom.kwanya}@gmail.com

Abstract. Universities in Kenya have invested in library services and resources to support their teaching and learning activities. However, these resources are not used adequately leading some universities to develop partnership programmes with students, known as knowledge ambassadors, to educate their peers about library resources and services available in their institutions. This study sought to describe the concept of knowledge ambassadors from a knowledge management perspective; investigate where and how it has been deployed in universities in Kenya; analyse how effective knowledge ambassador programmes in Kenyan universities have been; identify the challenges affecting its effective use for tacit knowledge creation and sharing; and propose strategies which can be used by academic institutions in Kenya to enhance the effectiveness of their knowledge ambassador programmes.

Primary data was collected through interviews with knowledge ambassadors and librarians selected through information-oriented purposive sampling from four public universities in Kenya. The findings indicate that eight private and public universities in Kenya have launched knowledge ambassador programmes. The benefits of the programmes include increased usage of the information resources by users in the institutions as well as deepened user participation in the design, development and deployment of library services and products. The findings also revealed that the impact of the programmes is hampered by time constraints; misconception about the expectations of the ambassadors; inadequate infrastructure; and inadequate funding. These findings can be used by universities in Kenya to develop policy frameworks which leverage the impact of knowledge ambassador programmes on tacit knowledge creation, validation, sharing and use.

Keywords: Tacit knowledge · Knowledge ambassadors · Knowledge peer groups · Knowledge volunteers · Kenya

1 Introduction

Tacit knowledge can be perceived as the personal, private, informal, un-codified, context-based experience or know-how [6, 14]. It also refers to the insightful and hard-to-define experiential knowledge. It is the most precious source of original new knowledge [13] and the basis of expertise which can give an organisation a unique com petitive advantage in any corporate environment [11, 17]. According to Botha *et al.* [5]

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 794–805, 2015.
DOI: 10.1007/978-3-319-21009-4_58

tacit knowledge is found in human minds and includes values, attitudes, skills and capabilities. It is generally considered to be non-verbal, inarticulable, unconscious or ineffable [1, 15, 16]. Tacit knowledge provides the foundation which enables individuals to deal with emerging issues [2, 10] and facilitates rapid decision making and action without requiring lengthy deliberation [9]. Atwood [3] explains that tacit knowledge includes personal contacts, networks and relationships which are difficult to remove from the owner.

Tacit knowledge is generally acquired through direct experience. However, as Dalkir [8] argues it can be transferred through coaching and mentoring on a one-to-one, face-to-face basis. O'Dell and Hubert [14] supports this view and explain that tacit knowledge is so hard to catalogue, difficult to document and ephemeral that it can only be shared through interactions. One of the major challenges hampering the effective creation, sharing and use of tacit knowledge is the difficulty of capturing, storing or sharing it. Dalkir [8] explains that highly skilled, experienced and expert individuals find it harder to explain their know-how. He explains further that novices, on the other hand, are more apt to easily verbalise what they are doing because they typically follow manuals or how-to processes. Therefore, tacit knowledge is generally created and shared through diverse forms of conversations and collaboration between individuals or groups.

As centres of scholarly endeavour, universities greatly depend on the effective creation, sharing and transformation of tacit knowledge in the form ideas, innovations, networks and linkages into explicit knowledge. Most universities commit huge resources to knowledge development through research and other related programmes. These efforts are, however, being held back by internal competition, lack of trust, inappropriate organisational structures and culture and lack of cooperation between knowledge teams. Consequently, many universities are currently exploring diverse strategies of mitigating poor tacit knowledge creation and sharing in the institutions. One of the strategies some universities have adopted is the use of students as knowledge ambassadors. The knowledge ambassadors are student volunteers who are willing and excited to promote information services to their peers. There is great need for peers to participate in knowledge sharing because it improves communication between information centres and peer community; develops user-centred services; and taps peer energy and creativity. According to Collison & Parcell [7] knowledge ambassador programmes involve peers learning from each other and trying to come up with solutions to their specific knowledge needs and thus create a highly focused environment for knowledge sharing and cooperation. Knowledge sharing is also done through mentoring which is like a corporate social responsibility programme amongst peer groups as they share their knowledge, experience and ideas with others. Knowledge ambassador programmes thrive on mutual commitment, respect and trust.

2 Study Rationale

Universities in Kenya have invested in library services and resources to support their teaching and learning activities. The resources include digital and electronic collections. However, library usage statistics indicate that these resources are not used adequately by the primary academic library users – the academic staff and students.

The low usage has been attributed to a lack of awareness of the resources; inadequate information literacy skills amongst the library users; presence of alternative sources of information in the academic infosphere; overloaded curricula leaving the students and staff with little or no time to use the library resources; and a poor perception of the real value of academic library services and resources. Most universities have deployed diverse marketing programmes in their efforts to increase the usage and impact of library resources and services. Some of these programmes include social media marketing, exhibitions, posters, promotional merchandise and information literacy classes. In spite of these efforts, most academic libraries in Kenya still register low usage. This situation led some universities to develop partnership programmes with students, known as knowledge ambassadors, to educate their peers about library resources and services available in their institutions.

Although knowledge ambassador programmes were launched over two years ago, no study seems to have been conducted to investigate their role in tacit knowledge creation and sharing. Since their actual or potential role in tacit knowledge management has not been investigated, most of the institutions deploying the programmes have not received the possible optimum benefits. Given the pressure on universities and academic libraries to remain relevant in a fast-changing and competitive world, the contribution of effective tacit knowledge management to their overall performance and impact cannot be ignored. This study sought to bridge this gap by describing the concept of knowledge ambassadors from a knowledge management perspective; investigating where and how it has been deployed in universities in Kenya; analysing how effective knowledge ambassador programmes in Kenyan universities have been; identifying the challenges affecting its effective use for tacit knowledge creation and sharing; and proposing strategies which can be used by academic institutions in Kenya to enhance the effectiveness of their knowledge ambassador programmes.

3 Theoretical Framework

This study was anchored on Max Boisot's i-space knowledge management model. In this model, Boisot proposes the concept of "information goods" which users extract from data but dependent on their expectations and prior knowledge. He argues that the effective movement of information goods is dependent on senders and receivers sharing the same coding scheme or language [8]. He further argues that effective extraction and sharing of information goods depends on the context and interpretations of the users [4]. This model underscores the role of shared coding scheme, language, context and interpretation in the creation, extraction or sharing of knowledge. Boisot's i-space is similar to Nonaka and Konno's [12] "ba" concept in the improved Nonaka and Takeuchi's [13] Socialisation, Externalisation, Combination and Internationalisation (SECI) model in which the role of a conducive space is emphasised as a determining factor of the effectiveness of key knowledge management processes such as origination, dialogue, systematisation and application [12]. Both models emphasise the role of face-to-face interaction in effective knowledge creation and sharing. Figure 1 below presents the i-Space model.

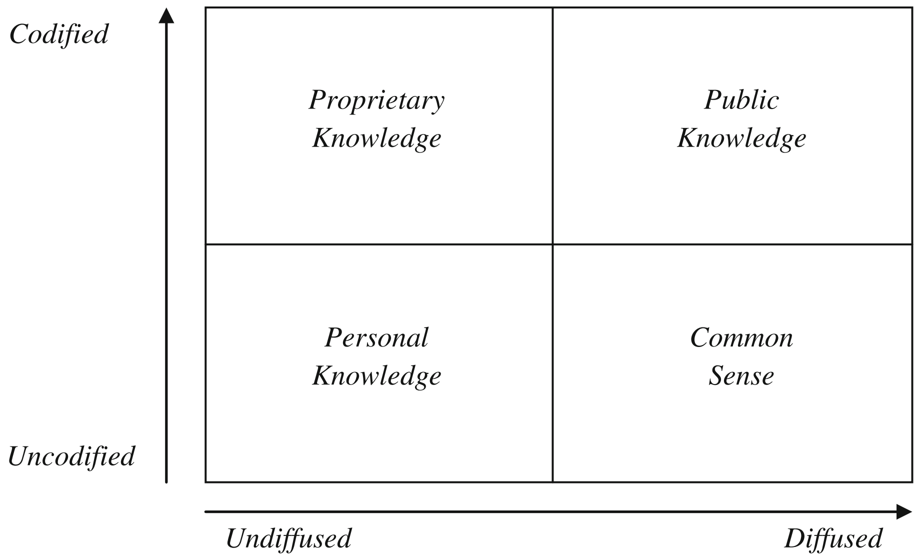


Fig. 1. i-Space model

The authors used Boisot's i-space model complemented by Nonaka and Konno's [12] "ba" to explore the role of socialisation and face-to-face interactions in tacit knowledge creation and sharing. They also used the two models to understand the attributes of a "ba" conducive for effective tacit knowledge management. Knowledge ambassadors as peers of their fellow students have a shared coding scheme (knowledge categorisation), language and context. Knowledge ambassadors require a "ba" which is safe, trustworthy, free, accessible, convenient, adaptable, natural, supportive and usable to be effective.

4 Methodology

This study was designed as an exploratory survey since the concept of knowledge ambassadors is fairly new in Kenyan universities. Exploratory research approach was flexible thus enabling the authors to investigate the research questions exhaustively. Primary data was collected through key informant interviews with twenty (20) knowledge ambassadors and two (2) librarians selected through information-oriented purposive sampling from four public and private universities in Kenya. The interviews, which were conducted using semi-structured guides administered by the authors in the respective university campuses of the respondents, enabled the authors to obtain in-depth data from persons who understood knowledge ambassador programmes well. The use of key informant interviews also enabled the authors to obtain diverse opinions and perspectives based on the respondents' experiences and contexts. The authors also managed to obtain in-depth responses by adopting a face-to-face approach which enabled them to clarify responses. Additional information on knowledge ambassadors as peer educators as well as the challenges facing university libraries in Kenya was collected from relevant documentary analysis. Information thus obtained triangulated the primary data and enhanced the validity of the findings. The data was analysed using descriptive statistics. This analysis technique was chosen because it enabled the authors to summarise, interpret and describe the data in a way that reveals their meanings within the context of the study.

5 Findings

The findings of the study are presented and discussed hereunder.

5.1 The Concept of Knowledge Ambassadors in Kenya

The respondents indicated that knowledge ambassadors in Kenyan universities are student volunteers who support academic librarians in their efforts to increase the use of library information resources. They further explained that these volunteers are energetic, passionate and enthusiastic about the role libraries play in academic settings. The ambassadors act as peer educators empowering and encouraging their colleagues to use library resources. Knowledge ambassadors are identified, recruited and trained by the librarians.

The respondents also explained that they first got to know of the knowledge ambassadors' concept through a workshop conducted by Emerald Insight at the Catholic University of Eastern Africa in 2013. The training was primarily aimed at enhancing the use of e-resources available in Kenyan academic libraries by the students. The training was attended by thirty (30) students drawn from The Cooperative University College of Kenya, The Technical University of Kenya, Jomo Kenyatta University of Agriculture and Technology and the Presbyterian University of East Africa. The participants were exposed to the Emerald e-resources available in their libraries as well as how best to use them. They were also asked to share the same knowledge with their colleagues. In this regard, they were given library promotional materials to distribute to their friends. The original participants in this workshop became known as Emerald Knowledge Ambassadors. Members of this pioneer group initiated knowledge ambassador programmes in their universities.

The findings also show that knowledge ambassadors in Kenyan universities act as knowledge champions, change agents, peer trainers as well as knowledge advocates and brokers. As volunteers, they perform these roles on a part-time basis. Furthermore, their influence is localised and generally restricted to their peer groups and specific university campuses only. The respondents explained that to perform their roles effectively, knowledge ambassadors need flexibility, goodwill, positive attitude, resilience, tolerance and good interpersonal skills. They also need to be proactive, analytical, committed, and familiar with their library information services and resources. They also need to earn the respect of their colleagues and librarians.

From the foregoing, it is evident that knowledge ambassadors are actually knowledge champions willing and excited to work with librarians to enhance the use of library resources. These ambassadors, just like other knowledge champions, need specialised technical skills and personality traits to succeed in their communities.

Librarians as well as the other stakeholders, such as scholarly publishers, support the ambassadors to develop the technical skills and aptitude to be effective champions. Although universities in Kenya and other countries operate in diverse contexts, the concept of knowledge ambassadors can be customised and replicated in any academic setting.

5.2 Deployment of Knowledge Ambassadors' Programmes in Kenya

The findings revealed that knowledge ambassadors in Kenyan universities are identified and recruited by librarians through diverse outreach programmes including training events, library or information weeks, exhibitions, and specialised information literacy programmes. During these programmes, the librarians appeal to the participants to volunteer their time and skills as knowledge ambassadors. The librarians take the contacts of those who express interest and invite them to capacity development programmes through which they are empowered to work as knowledge ambassadors.

The respondents said the concept is currently deployed by academic libraries at The Technical University of Kenya, Jomo Kenyatta University of Agriculture and Technology, Strathmore University, Cooperative University College of Kenya, Daystar University, Presbyterian University of East Africa, Tangaza University College and Catholic University of Eastern Africa. It is noteworthy that all the universities implementing knowledge ambassador programmes, except three, are privately owned. Recognising the fact that there are more chartered public (22) than private (14) universities in Kenya, the higher number of private universities deploying knowledge ambassador programmes may be explained by their customer-centric orientation. Given that private universities operate in more competitive environments than their public counterparts, they may have been encouraged to embrace knowledge ambassador programmes as a means of enhancing their competitive advantage. Furthermore, most private universities in Kenya are smaller than the public universities and are keen to integrate the students' participation in as many programmes as possible. This situation may have also been motivated by the strict accountability frameworks private universities operate in. Academic libraries in private universities are, for instance, required to account for their resources as a way of demonstrating the return on investment they accrue for their parent institutions. With the increasing demand for better impact by public universities through performance contracts, it is expected that more of them will adopt knowledge ambassador programmes as one of the strategies to enhance their relevance. Similarly, with the growing popularity of university ranking frameworks such as Webometrics, more universities are likely to initiate knowledge ambassador programmes to increase their research output thereby enhancing their ranking.

5.3 What Knowledge Ambassadors in Kenyan Universities Do

Asked to indicate what knowledge ambassadors in their universities do, the respondents said that they generally promote library resources and services. They pointed out that the ambassadors particularly encourage the use of electronic resources the universities have. These electronic resources are made available by the libraries by subscribing to e-journals available through the Kenya Library and Information Service

Consortium (KLISC) in collaboration with the International Network for the Availability of Scientific Publications (INASP). The ambassadors use myriad tactics in their work. For instance, in one university, the knowledge ambassadors organised a beauty pageant dubbed “Mr and Miss Open Access” to promote the use of electronic open access materials.

The respondents still emphasised that the knowledge ambassadors also help in promoting the other library resources as well. They do so by managing social media platforms for their libraries, holding exhibitions, developing and promoting library websites, and designing and disseminating library brochures. Respondents from one private university also reported that knowledge ambassadors also promote library services and resources using the university’s FM radio station so as to reach more actual and potential information users. Still in another university, the respondents reported that the knowledge ambassadors participate in corporate social responsibility (CSR) programmes. The example of such a CSR programme is supporting the services offered by a community library in a slum area in Nairobi. Apart from helping to offer general library services, the ambassadors also tell stories or read books for pupils using the facility.

In two universities, the knowledge ambassadors also organise learning events such as knowledge cafes where the students meet to share experiences on topical issues of interest. These meetings, generally held once a semester, provide platforms where students exchange tacit knowledge and create contacts for subsequent interactions and learning. The respondents explained that although the forums were originally meant for learning about library services and resources, they have since mutated to cover all knowledge areas of interest to the students. Consequently, they are used for brainstorming, mentorship, incubation and innovation. Their informal and voluntary nature has endeared them to many students. In fact, they are gradually turning into communities of practice which are better structured and focused. Given that modern students generally trust their peers more than authorities, these forums provide platforms with a greater potential to facilitate effective tacit knowledge creation and sharing.

Many of the students in universities in Kenya are in their late teens or early twenties. This age group struggle with many social challenges which in some cases result in dropouts, poor performance or deviant behaviour. Since most knowledge ambassadors are generally disciplined, diligent and cooperative students, they act as role models for their colleagues who learn from their experiences and determination thus reducing the impact of teen challenges on their studies. Anecdotal evidence from the respondents seems to indicate that knowledge ambassadors are excellent academic performers who are respected by their peers. Thus their influence goes beyond the original scope of work.

Knowledge ambassadors also help to extend library support services to the students’ halls of residence and other places of presence where they act as embedded librarians. Given their first-hand understanding of their curricular requirements as well as their experience with library resources, they are able to identify and respond to their colleagues’ information needs promptly and more accurately. Academic librarians are not able to be available in all places on campus or where the students live or learn. Conversely, the knowledge ambassadors walk, work, live and learn with their colleagues and are able to offer prompt support at the point of need. However, the respondents emphasised that the ambassadors need a higher level of information literacy to be able to help their colleagues competently. Nonetheless, all the librarians who responded to the study asserted that knowledge ambassadors go through elaborate training which builds their capacity to help their colleagues as a first-aid measure where and when it is not possible to get services from the librarians.

The knowledge ambassadors have also conducted workshops on plagiarism and related issues; competitions on issues related to information services and resources; compiled and disseminated documentaries on relevant topics; written and circulated poems; developed scripts and acted skits and short plays; as well as developed and presented conference papers on issues which influence the access and use of diverse information resources in academic libraries.

It is evident from the foregoing that knowledge ambassador programmes in Kenyan universities were originally conceived to support the promotion of electronic resources in their libraries. However, as the programmes matured, the ambassadors have gradually picked up other roles which are not directly related to libraries. Although the roles vary from institution to institution, they are generally aimed at providing the students with the knowledge and resources to enable them to make the best use of their time and opportunities on campus. Significantly, the ambassador programmes are gradually mutating into communities of practice and other forms of knowledge networks which provide "ba" conducive for tacit knowledge creation, validation, sharing and application. It is possible that the ambassador programmes will become better structured over time and be able to generate greater impact for their parent institutions.

5.4 The Benefits of Knowledge Ambassadors Programme

All the respondents were of the view that the knowledge ambassador programmes have generated benefits in their institutions. They identified the benefits as increased usage of the information resources by users in the institutions; better attitudes and appreciation for the librarians and library resources; deepened user participation in the design, development and deployment of library services and products; improved information literacy skills amongst the students; increased number and scope of innovations amongst the students and other library users; better understanding of the challenges, needs and wants of students by the librarians; reduction of operational costs associated with the promotion of library services and resources; greater ownership of libraries by the students; increased relevance of library services and resources; and increased effectiveness of the academic libraries in fulfilling their mandates to their parent institutions.

The findings above demonstrate the fact that knowledge ambassador programmes have real benefits for academic institutions. The benefits above, albeit varied, mainstream the impact of libraries on organisational performance. Librarians are encouraged to harness the full potential of these programmes in their contexts.

5.5 The Challenges Hampering the Impact of Knowledge Ambassadors

Although the respondents were unanimous that their institutions have benefited from knowledge ambassador programmes, they also pointed out that the institutions are not realising the full potential of the programmes. All the respondents said that time constraint was the major challenge the knowledge ambassadors face. Given that they are fulltime students who must also carry full academic loads, most of them find it difficult to create adequate time for ambassadorial activities. Consequently, some programmes have had to be delayed or conducted over a longer period of time than is desirable.

Another major challenge is a misconception of what the knowledge ambassador concept really is. This misunderstanding results in the ambassadors taking up roles they are either not able or not expected to perform. The misunderstanding sometimes results in conflicts or poor delivery of services. One university also reported that some of the ambassadors used their positions as a springboard into student politics or to solicit other favours from the librarians, students and the university management. Such ambassadors were not effective because they were perceived as dishonest.

The programme has also been affected by its volunteerism nature. It is not easy for the librarians to demand results from the ambassadors knowing that they are just volunteering. Similarly, some ambassadors do not feel obligated to perform their roles and only participate in the programmes which they find exciting or gratifying. This situation was identified by the librarians as one the major risks of the programme. They have attempted to mitigate it by making the programme as exciting as possible and giving the ambassadors incentives such as branded merchandise and opportunity to attend training workshops.

The other challenges identified by the respondents include lack of adequate funds to facilitate outreach activities; inadequate infrastructure such as computers and bandwidth to support the effective access and use of electronic resources amongst the students; overstretching the programme to include issues which are not related to information access and use; burnout of the ambassadors; unpredictable information needs and seeking behaviour; and competing student, library and institutional programmes.

5.6 Strategies Institutions Can Use to Enhance the Impact of Knowledge Ambassadors

The respondents recommended that for the programme to run efficiently and effectively, the institutions should allocate adequate funds to it. The librarians should also develop and implement an extensive mentorship programme to the ambassadors so as to fully build their capacity to perform their roles. The respondents also emphasised that the libraries should develop a sustainable framework for appropriately rewarding the ambassadors, especially those who excel in their duties. This can be done by giving incentives like participation certificates and publishers tokens. Furthermore, the institutions can treat knowledge ambassador programme as student work programmes through which the ambassadors may be paid a stipend. This way, the libraries can demand accountability from them. The other strategies recommended by the respondents include more investment in infrastructural development; closer but realistic monitoring and evaluation of the programme activities; identifying and maintaining focus on the major issues only so as to ensure the resources are not stretched too thin to make impact; and structuring the ambassador recruitment process to ensure that only those who meet the essential requirements are engaged. The librarians should also take time to explain the essence of knowledge ambassadorship to ensure a common understanding of what it entails.

6 Conclusion

The findings of the study reveal that knowledge ambassador programmes provide a "ba" conducive for tacit knowledge creation, validation, sharing and use in Kenyan universities. However, their impact is hampered by ideological, infrastructural and logistical challenges. The universities need to address these challenges using the recommendations above so as to leverage the impact of the programmes in facilitating the creation of unique tacit knowledge to increase their competitive advantage.

7 Implications of the Findings of the Study

The findings of this study can be used by universities to develop policies and action plans which leverage the operations and impact of knowledge ambassador programmes. The findings also underscore the value of peer education programmes in tacit knowledge management and can be used by librarians to develop information environments which are user-centric and promote collaboration in providing solutions to challenges hampering their impact. The findings may also be used by knowledge management professionals to develop theories and models on which knowledge ambassador programmes can be anchored.

The Interview Guide

1. Describe the origins and concept of knowledge ambassadors

----

2. How does the concept work in Kenya?

----

3. Where has the concept been implemented in Kenya?

----

4. What benefits have institutions implementing the concept accrued?

----

5. What challenges hamper the effective use of the concept in Kenya?

----

6. What strategies can academic institutions in Kenya use to enhance the effectiveness of knowledge ambassadors?

----

7. Any other issue you’d like to add?

References

1. Ambrosini, V., Bowman, C.: Tacit knowledge: some suggestions for operationalization. J. Manage. Stud. **38**(6), 811–829 (2001)
2. Argyris, C.: Tacit knowledge and management. In: Sternberg, R.J., Horvath, J.A. (eds.) Tacit Knowledge in Professional Practice, pp. 123–140. Lawrence Erlbaum Associates, Mahwah (1999)
3. Atwood, C.G.: Knowledge Management Basics. ASTD Press, Alexandria (2009)
4. Boisot, M.: Knowledge Assets. Oxford University Press, Oxford (1998)
5. Botha, A., Kourie, D., Snyman, R.: Coping with Continuous Change in the Business Environment, Knowledge Management and Knowledge Management Technology. Chandos Publishing Ltd., Oxford (2008)
6. Brown, J.S., Duguid, P.: Organizing knowledge. Calif. Manag. Rev. **40**(3), 90–111 (1998)
7. Collison, C., Parcell, G.: Learning to Fly: Practical Knowledge Management from Leading and Learning Organizations. Capstone Publishing, Chichester (2004)
8. Dalkir, K.: Knowledge in Theory and Practice. The MIT Press, Cambridge (2011)
9. Herbig, B., Bussing, A., Ewert, T.: The role of tacit knowledge in the work context of nursing. J. Adv. Nurs. **34**(5), 687–695 (2001)
10. Horvath, J.A., Forsythe, G.B., Bullis, R.C., Sweeny, P.J., Williams, W.M., McNally, J.A., Wattendorff, J.A., Sternberg, R.J.: Experience, knowledge and military leadership. In: Sternberg, R.J., Horvath, J.A. (eds.) Tacit Knowledge in Professional Practice, pp. 39–58. Lawrence Erlbaum Associates, Mahwah (1999)
11. Marwick, A.D.: Knowledge management technology. IBM Syst. J. **40**(4), 814–830 (2001)
12. Nonaka, I., Konno, N.: The concept of “ba”: building a foundation for knowledge creation. Calif. Manag. Rev. **40**(3), 40–54 (1998)
13. Nonaka, I., Takeuchi, H.: The Knowledge Creating Company. Oxford University Press, Oxford (1995)
14. O’Dell, C., Hubert, C.: The New Edge in Knowledge: How Knowledge Management Changes the Way We do Business. Wiley, Hoboken (2011)
15. Patel, V.L., Arocha, J.F., Kaufman, D.R.: Expertise and tacit knowledge in medicine. In: Sternberg, R.J., Horvath, J.A. (eds.) Tacit Knowledge in Professional Practice, pp. 75–100. Lawrence Erlbaum Associates, Mahwah (1999)
16. Tsoukas, H.: Do we really understand tacit knowledge. In: Easterby-Smith, M., Lyles, M.A. (eds.) The Blackwell Handbook of Organizational Learning and Knowledge Management, pp. 410–427. Blackwell Publishing Ltd., Malden (2003)
17. Wellman, J.L.: Organizational Learning: How Companies and Institutions Manage and Apply Knowledge. Palgrave Macmillian, London (2009)

An Analysis of the Content of Knowledge Management-Related Courses

Mitja Gradišnik(✉), Vili Podgorelec, and Marjan Heričko

Faculty of Electrical Engineering and Computer Science,
University of Maribor, Smetanova ulica 17, 2000 Maribor, Slovenia
{Mitja.Gradisnik, Vili.Podgorelec,
Marjan.Hericko}@um.si

Abstract. Operating in a knowledge-based economy requires a wide range of additional skills and competences. An important part in the transfer of new skills and competences to the labor marker influence the design of courses offered at higher education institutions. To support students' professional career development, the content of the courses should constantly be aligned with new skills and competences required by the economy. In this paper we will present our research on the content of knowledge management-related courses. During our research we gathered a list of knowledge topics from the domain of knowledge management introduced within knowledge management-related courses, which enjoy greater support in the academic sphere at engineering-oriented higher education institutions. This research looked at the content of 41 knowledge management-related courses offered at 79 higher education institutions in 21 countries in the European Union.

Keywords: Knowledge management · Higher education · Teaching knowledge management

1 Introduction

Since the 1990 s there has been an obvious shift from an information-based economy to a knowledge-based economy [9]. The knowledge-based economy has been marked by a completive business environment that demands constant advances and innovations in the development process of new products and services. The success of businesses operating in an increasingly competitive marketplace of the knowledge-based economy depends critically on the quality of knowledge which these organizations apply to their key business processes [14]. Therefore, the creation, management, and sharing of knowledge within the organization have become an important factor behind competitiveness that should not be overlooked by companies.

The emergence of the knowledge age has brought with it significant responsibilities with regard to the need to adjust our sociological and economic structures to provide opportunities for improvement and growth [16].

For modern businesses it is crucial that they "know what they know" [14]. To adopt to the fast-changing economic environment and to ensure its long-term business stability, businesses operating in a knowledge-centric economy require a whole range of

© Springer International Publishing Switzerland 2015
L. Uden et al. (Eds.): KMO 2015, LNBIP 224, pp. 806–818, 2015.
DOI: 10.1007/978-3-319-21009-4_59

new skills, competences and practices. Knowledge management can be seen as a process of transforming information and intellectual assets into an enduring value. In this process, businesses are concerned with identifying and formalizing existing knowledge, acquiring new knowledge for future use, achieving it in organizational memories and creating systems that enable effective and efficient application of knowledge within the organization [14]. From an operational perspective, knowledge is used in everyday practice by professional personnel, who need access to the right knowledge at the right time.

Higher education institutions are undoubtedly an important link in the chain of providing suitably trained workers equipped with all the necessary skills and competences. The primary goal of higher education is to make available a selection of the virtue, intellectual and technical capital to meet the needs of the society it serves [9]. As the core tool in meeting the intended learning outcomes, curriculum is the key factor to bringing changes to graduates, who are future knowledge workers [9]. For higher education institutions it is therefore crucial to stay in sync with the latest demands of the labor market, so that the latest changes in the labor market can be incorporated into curriculum subjects as quickly as possible.

Over the last few years knowledge management has been growing in popularity in the academic sphere. The continued drive for improvements in higher education encourages researchers to address the gap between existing study programs and the need for knowledge-skilled professions demanded by the present-day labor market.

How to include topics of knowledge management in the curriculum of higher education institutions has been the subject of debate in many studies [1, 2, 3]. Al-Hawamdeh [1] and Bedford [3] addressed the process of establishing a graduate-level program of knowledge management at a university. With the newly established graduate-level study program they tried to provide an answer for the need for an emerging new profession in today's knowledge society and economics. Argamon et al. [2] described a process of extending the curriculum of the undergraduate program of Computer Science. The authors introduced five additional courses to the study program that cover perspective disciplines of knowledge management. With the additional knowledge management-oriented courses they tried to reach a deeper understanding of techniques for information and knowledge management systems among graduates.

The listed surveys mainly address the process of establishing knowledge management-related study programs and courses at higher educational institutions and do not go into details about the course topics selection activities. During the literature review, we did not encounter any survey that directly analyzed topics taught within knowledge management-related courses.

In our study we intended to research the inclusion of IT support systems in the syllabus of knowledge management-related courses. By its nature, knowledge management is a multidisciplinary domain. Our primary focus of study is the technological perspective of knowledge management. In the study we would like to identify topics within the domain of knowledge management, which are the most frequently included in the content of knowledge management-related courses. We are restricted to courses offered within study programs at universities. The primary focus of the study is to identify the topics related to IT support systems of knowledge management processes and practices. In the study we would like to restrict ourselves to the common economic space of the European Union.

Based on these presuppositions we would like to answer the following two research questions. First, we would like to identify which topics from the domain knowledge management IT support are offered within KM-related courses at universities. For each identified topic, we would like to determine the frequency of occurrences within courses. Secondly, we would like to explore which domains of knowledge management are covered by the previously identified topics.

The results of our research would provide us with solid insight into what kind of technological skill and competences today's graduates possess in the domain of knowledge management. The results can be also be a good starting point for the eventual renovation of knowledge management-related courses.

This paper is structured as follows. The introduction establishes the scope and purpose of the paper and gives the necessary background information to support. Following the introduction, a section explains the need for knowledge management and introduces its main disciplines. We then describe the method for empirical data collection and analysis in Sect. 4. Section 5 outlines the findings on our research. Finally, in Sect. 6, we summarize our research and give concluding remarks.

2 Knowledge Management

Although the term knowledge management started to receive considerable attention in late 1980 s in both the academic and economic world, knowledge management as a practice has been around for many decades. It could also be argued that knowledge management has been around for far longer than the actual term has been in use [9]. Throughout history, people have invented different approaches to allow them to pass knowledge on from one generation to the next or to share it in order to build on earlier experiences. Modern concepts and practices of knowledge management have evolved over the last two decades, in times when knowledge was recognized as a critical resource vital to continuing economic growth.

Regardless of the technological end economic progress over the decades, knowledge stays human-centered. Knowledge management can be seen as a process of creating, collecting, processing, storing and diffusing knowledge. The aim of knowledge management activities is achieving better outcomes in decision-making capabilities, speeding up business processes, reusing best practices and reducing unnecessary work.

Although organizational knowledge is mostly seen as an intellectual asset or resource, it possesses characteristics which fundamentally distinguish it from other valuable assets in an organization – it cannot be easily stored and protected. The large part of an organizational knowledge asset is comprised by employees. One of the main aims of knowledge management is to ensure that these valuable knowledge assets are not lost for the organization through retirement or turnover of personnel. With the aim of preserving organizational intellectual assets, the tacit knowledge held by personnel should be transformed during knowledge management processes into internal explicit knowledge such as reports, guidelines, course syllabi, theses and databases or shared among employees [10].

It is hard to find a single generally accepted definition of knowledge management. One of the reasons for this is the fact that knowledge management is a relatively new field and the term remains a topic of further debate. According to the above-stated, knowledge management could be defined as: *the deliberate and systematic coordination of an organization's people, technology, processes, and organizational structure in order to add value through reuse and innovation. This coordination is achieved by creating, sharing, and applying knowledge as well as through feeding the valuable lessons learned and best practices into corporate memory in order to foster continued organizational learning [9].*

According to Daklir [9] at the moment there are over 100 published definitions of knowledge management, and most of them could be considered appropriate. The diversity of published definitions can be explained by the fact that each of the main concepts of knowledge management, people, technology and processes, represent an independent perspective on knowledge management. The objectives of knowledge management are achieved by the mutual interlacing of competences from a vast number of diverse fields. It is also important to emphasize that knowledge management cannot be a responsibility of a selected group of people or selected department within an organization. It is a process that should include the majority of the personnel in an organization.

2.1 The Multidisciplinary Nature of Knowledge Management

It is clear that knowledge management is not about IT or any other single discipline. It is the amalgamation of technology, processes, and people, which augments knowledge creation, retention, and transfer [1]. In general, the interdisciplinary nature of knowledge management requires a broad range of skills and competences that need to be acquired. Efficient knowledge management is based on domains such as library and information sciences, linguistics, process engineering, sociology, organizational theory, and education and training [9]. Although the list of domains [11] involved in knowledge management activities is quite extensive and diverse, we can identify three distinct perspectives on knowledge management. Each of the listed perspectives has its own focus on the operation and objectives, measured by critical success factors.

2.2 Organizational Perspective

People, not systems, manage knowledge [15]. The organizational perspective is focused in an engagement of organizational structures with the aim of the effective realization of business goals. The organizational structures can be described as communities or groups of people with common interests and experiences. This group of people will often work on similar problems inside an organization. To support the efficiency of personnel, organizations can enforce policies and practices that encourage the sharing and diffusion of knowledge among personnel. Knowledge management activities on an organizational level are built on collegiality and teamwork. Employees are encouraged to share with others what they know and what they are learning.

They are encouraged to build common repositories of resources and artefacts, which can be a good basis for continuing the learning process and prevention of knowledge loss due to a fluctuation in personnel.

The activities of knowledge management are not limited to just the borders of an organization. In that case, we can talk about interorganizational sharing and the exchange of knowledge. Knowledge obtained from external sources can be in an explicit or tacit form. External explicit knowledge sources are considered artefacts carried by different types of media. External tacit knowledge can be provided by a variety of external domain experts included in the business process of an organization.

2.3 Business Perspective

From a business perspective, knowledge management activities are mostly concerned with protecting and exploiting an organization's knowledge or intellectual assets to produce revenue streams [15]. In a knowledge-based economy, the intellectual capital possessed by a company is a key factor in its long-term development. How importunate knowledge or intellectual assets are, is indicated by the fact that the value of intellectual capital represents a significant share in the market value of a company. From a business perspective, the activities of knowledge management are primarily concerned with managing intellectual assets in a manner that encourages knowledge codification and stores these artefact registers in content repositories. From a business perspective, knowledge management should protect knowledge and intellectual assets from being lost due to employee turnover or for any other reason.

2.4 Technological Perspective

Even though technology is not the frontal actor of knowledge management, it plays an important role in its efforts of fulfilling business goals. One of the aims from a technological perspective is how to take advantage of technologies as a core organizational resource [15]. Within the domain of knowledge management, technological perspective addresses the challenges of using information technologies with the aim of supporting knowledge management processes and practices within organizations. Technological support consists of a vast range of information systems, tools, and related technologies, which are primarily intended to support personnel by their business efforts, codifying knowledge and creating new knowledge by analyzing a vast amount of data. Therefore, knowledge management from a technological perspective can be understood as the concept under which information is turned into actionable knowledge and made available effortlessly in a usable form to people who can apply it [9].

The list of technologies and tools involved in the knowledge management process is extensive and covers a variety of domains within computer science. The core fields of emerging subdisciplines of knowledge management are text analysis, data mining, information retrieval, and database systems [2].

3 Research Approach

In order to answer our research questions we gathered the topics of syllabus of knowledge management-related courses in the first stage of the research and then classified them in the second stage. The primary aim of the classification of the identified topics is to see a distribution of the topics within the spectrum of knowledge management technologies. The spectrum provides us the insight into which processes and practices of domain of knowledge management enjoy the greatest support, according to knowledge passed to the students.

Some higher education institutions introduce students with knowledge management topics earlier on during the studies, during the first degree studies, while other institutions, who consider this topic more advanced, introduce students to these topics during their master's degree studies. For this reason we collected and analyzed data separately for the first degree and for the master's degree study programs.

3.1 Identification of Topics

The search for knowledge management related subjects, included in our research, was based on a list of universities[1] that signed a bilateral agreement on student exchanges with our faculty as part of the Erasmus student exchange program.[2] As part of the Erasmus student exchange program, our students can choose between 104 different partner institutions in 21 countries across the European Union. In addition to other types of education institutions (e.g. institutes, academies) there are 79 universities in our student exchange network. In our research we also included courses offered at universities, which were also in the focus of our interest. In order to allow students to carry out part of their studies outside of their home university, study plans and courses at universities within the Erasmus network are compatible and structurally similar. The structural similarity of study programs, which is mostly the result of the unification of higher education curricula as part of the Bologna Process, helped us compare the courses and content passed on to students within these courses. This is also one of the reasons why others types of education institutions besides universities were not included in our research. The list of partner universities also includes some universities located in non-EU countries, study plans of which were also not included in our research.

The content of courses used in this research was obtained by manually reviewing study plans and corresponding courses descriptions available on the public web sites of individual education institutions from the list. In this stage of our research, the content of study programs was analyzed by only one rater. In order to reduce subjectivity it would be reasonable to include additional raters in the future.

Course descriptions usually consist of title, lecture, number of ECTS (or country specific equivalent) points and a list of topics provided within the course. The topics have usually been specified in the form of concise text or a list of lectured topics.

---

1 http://feri.um.si/site/assets/files/1413/iia_feri_2_outgoing_13_1_2015.pdf.

2 http://feri.um.si/studij/erasmus/.

In our research we collected detailed descriptions of university courses related to knowledge management. In our research we only included courses that directly addressed topics of knowledge management. A course was considered as relevant for our study if (1) knowledge management was recognized from its title or (2) any of the domains of knowledge management [11] were recognizable from the course title and knowledge management was one of topics discussed within the course. The subjects that address knowledge management topics indirectly (for example, address approaches or technologies that could be used to support knowledge management and are introduced to students in the context of domain of knowledge management domain) were ignored. The example of such technologies are artificial neural networks or semantic web technologies. Both of these technologies often appear in courses as an enabling technology for knowledge management systems or tools but knowledge management is not the only domain where mentioned technologies can be used. If the connection between topic and knowledge management was not clearly evident from the description of the course, the subject was not included during further analysis.

For the each subject that met our conditions we tried to identify the following parameters: (1) the title of a subject, (2) the academic year in which the subject was offered, (3) the kind of content that was passed to students as part of the subject.

Additionally, we kept a record of the number of instances of each identified topic. The number of instances helped us to identify which topics appear in knowledge management related subjects at higher frequencies.

3.2 Classification of Topics in Accordance with the KM Spectrum

In the second stage of our research we identified how much attention each separate subdomain of knowledge management gets from the perspective of amount of knowledge delivered to the students during the classes. To achieve that, we classified previously identified topics into groupings by separate subdomains of knowledge management. The classification of the topics was conducted based on knowledge management spectrum proposed by Binney [4]. The proposed spectrum is suitable in our study because it maps specific elements of IT support of knowledge management into the subdomains of knowledge management processes and practices. In short, the proposed framework enables the classification of knowledge management applications and technologies, which are used to support knowledge management.

The base layer of the spectrum consists of knowledge management applications which can be classified into six common categories (elements) of the spectrum. The elements of knowledge management spectrum are labeled (1) transactional KM, (2) analytical KM, (3) asset management KM, (4) process based KM, (5) developmental KM, and (6) innovation/creation KM [4] (Table 1).

In addition to the base layer of applications of a specific spectrum element, the knowledge management spectrum contains an additional layer consisting of enabling technologies, which are required for the implementation of knowledge management applications in a base layer. The author of the spectrum considers the term enabling technologies in a fairly broad way. Therefore, the term “enabling technology” can also encompass any technology, approach or subsystem that is crucial in the development process of a knowledge management application or tool.

Table 1. Descriptions of elements of the knowledge management spectrum

Element of KM spectrum	Description
Transactional	Knowledge is embedded (encoded) in the application of technology.
Analytical	Knowledge is created by analyzing a vast amount of data of information in order to derive trends and patterns.
Asset Management	Manage processes of knowledge management associated with the management of knowledge assets.
Processes	Cover the codification and improvement process, procedures and methods.
Development	Focus on increasing the competences or capabilities of an organization's knowledge workers.
Innovation/Creation	Encourage collaboration in the knowledge creation process of knowledge workers.

4 Research Results

In this research we reviewed the curricula of partner universities at our faculty's Erasmus student exchange network, identified knowledge management-related courses and analyzed its content description in order to identify knowledge management topics taught within these courses.

4.1 Identified Topics Within KM Courses

Despite the large number of examined course descriptions (the curricula of 79 universities in the domain of computer science or informatics) we included in our analysis a relatively small number of collected content descriptions of knowledge management-related courses. One of the reasons for the relative small amount of collected material from the course content description is that many universities in their study programs do not offer any course that directly addresses topics from the knowledge management domain. It turned out that the knowledge management-related courses more often appear in the curricula of study programs in the domain of Information Technologies or Informatics than in study programs in the domain of Computer Science. The second reason for the relatively small number of collected course descriptions was the unavailability of a detailed content description taught within a course. Many education institutions do not prepare English versions of their course description documents nor publish these documents on their public web sites. Many of the analyzed descriptions were written in a generic way, so that the topics taught within the courses could not be identified. In our content analysis we included 41 course descriptions. 17 course descriptions were identified and analyzed in the first degree study programs and 24 descriptions were found within master's degree study programs.

Within each analyzed content description of knowledge management course we identified a number of key course topics that can directly support knowledge management processes and practices from the perspective of IT (technological perspective).

During the analysis we identified 17 main topics in the first degree study programs and 17 main topics offered within the knowledge management-related subjects in the master's degree study programs. The list of identified topics was used in the next stage of our research, in which the topics were classified according to the knowledge management spectrum.

During the content analysis we identified 6 course descriptions, in which we could not identify any of the topics from our search domain. Four of them were identified within first degree study programs and two of them within the master's degree study programs. The courses offer a quite general view on knowledge management from the organizational perspective and do not give any special attention to IT support of knowledge management processes and practices. Some of these 6 courses also shed light on the topics related to the cultural perspective of knowledge management and the importance of managing intellectual property.

4.2 Knowledge Management Spectrum

Finally, we classified the previously identified topics in accordance with the proposed knowledge management spectrum. The final results of our research are shown in Tables 2 and 3. For each classified topic we appended the number (the values in brackets after the title of topics) of occurrences of a specific topic during the subject description analysis.

Table 2. Topics of knowledge management-related courses (first degree level)

Element of KM spectrum	KM Applications	KM Enabling Technologies
Transactional	Recommender systems (1)	Semantic web (7)
	Planning systems (2)	Frame-based systems (1)
		Rule-based systems (4)
		Expert systems (2)
Analytical	–	Artificial neural networks (2)
		Decision trees (2)
		Agent systems (4)
		Text mining (2)
		Data mining (1)
Asset Management	Content management systems (1)	Search approaches (2)
Processes	–	Information management and workflows (1)
Development	Learning and training (1)	–
Innovation/Creation	Communities (1)	Web 2.0 (1)

In Table 2 it can be seen that transactional and analytical IT support systems are the focus of knowledge management related courses offered as part of undergraduate study

Table 3. Topics of knowledge management-related courses (master degree level)

Element of KM spectrum	KM Applications	KM Enabling Technologies
Transactional	Semantic-based apps (1)	Semantic web (10)
	Planning systems (1)	Rule-based systems (4)
		Frame-based systems (2)
		Expert Systems (3)
Analytical	Data warehouse systems (1)	Data mining (7)
		Text mining (1)
Asset Management	Document management systems (2)	Search approaches(3)
	Semantic knowledge bases (1)	
Processes	–	Information management and workflows (1)
Development	Learning and training (1)	–
Innovation/Creation	Communities (1)	Web 2.0 (1)
		Collective intelligence (1)

programs. The domain of transactional knowledge management systems, which deal with knowledge representation and codification, is mainly provided to students through an engineering process of knowledge based solutions based on semantic web technologies, rule-based systems and expert systems. The domain of analytical systems is covered via topics related to artificial neural networks, decision trees, agent systems, and text mining. Applications or enabling technologies intended to support other domains of knowledge management spectrum have a relatively low share of representation.

It is also evident that much more attention has been dedicated to topics related to enabling technologies or approaches required in an engineering process in knowledge management systems. Topics directly related to IT support applications or systems of knowledge management get significant less attention. During the analysis of content descriptions we identify recommender systems, planning systems and content management systems.

Table 3 shows identified topics of courses offered within the master's degree program. The topics are classified to accordance with the proposed knowledge management spectrum. In Table 3 it can be seen that transactional and analytical systems are the focus of knowledge management related courses at a master's degree level. Semantic web technologies, rule-based systems, and expert systems are at the forefront in the domain of transactional IT solutions, whereas data mining approaches and tools are the focus of the engineering process of analytical systems. Applications or enabling technologies intended to support others domains of the knowledge management spectrum have a relatively low share of representation, similar to the case of courses in first degree studies.

In comparison with the first degree study courses, the master's degree study courses introduce students with a wider range of knowledge management applications. Besides planning systems, document (content) management systems, and sharing solutions, students are additionally introduced to a data warehousing system, semantic knowledge bases and semantic applications, and software solutions intended to support the learning and training processes.

5 Discussion

When comparing the topics offered within the courses of undergraduate and graduate degree study programs we notice that there is only a slight difference between both levels of studies. The analyzed content of KM subjects indicates that the reviewed curricula in the master's degree study includes slightly more courses with a knowledge management related content. Also the diversity of topics offered within these courses is higher in master's degree studies. The differences probably stem from the fact that during the first degree studies, students are introduced with fundamental methodologies, approaches and technologies. Students at the master's degree study are introduced with more complex topics. When comparing the content passed to the students we notice that there is no major difference between both levels of studies. In the segment of analytical knowledge management the analyzed courses within the first degree study programs offer a wider range of enabling technologies, whereas the knowledge management subjects within the master's degree studies emphasize data mining approaches and technologies.

As shown in Tables 2 and 3, there is a much bigger emphasis on enabling technologies for knowledge management solutions, then introducing the final solutions themselves. An emphasis on enabling technologies can be perceived within courses of both first and master's degree studies. Within the analyzed knowledge management-related courses, students are primarily introduced to the enabling technologies (including methodologies and approaches) required to develop real-world knowledge management solution. Within some courses, students are introduced to the uses of different systems (e.g. Content Management Systems, Data Warehouse Systems) or open semantic knowledge bases (e.g. LinkedData, Wortnet).

According to the knowledge management spectrum used in our research, the degree of codification of knowledge moves from top to bottom across categories of the spectrum. The highest degree of knowledge codification includes transactional systems. Knowledge management systems devoted to the support of innovation/creation on the other side of the knowledge management spectrum focuses more on connecting people and consequently encourages the flow of tacit knowledge across organizations [4]. The distribution of identified topics across the knowledge management spectrum shows that topics classified on top of the spectrum (transactional and analytical KM) are apparently emphasized much more than topics from the bottom of the spectrum. Therefore, processes and practices related to knowledge representation and knowledge discovery are covered in more detail, while IT support for processes and practices related to and training knowledge sharing among employees is not in the focus of knowledge management teaching at universities.

6 Conclusions

As stated in the introduction, with the research presented within this paper we wanted to identify currently attractive topics offered within the domain of knowledge management courses at higher education institutions. The research was conducted on a sample of related universities in the Erasmus student exchange program. Although the structural similarity of study programs in the compared universities enables a better comparison, the affinity of universities in the sample can affect the generality of the results, which could be one of the limitations of this research.

From the research we can conclude that the topics learned within knowledge management related courses are rather heavily technology-oriented. The main focus is directed towards a variety of approaches and enabling technologies within an engineering process of knowledge management software solutions used to codify and discover knowledge.

In this manner, both research questions from the introduction were adequately answered. Firstly, we identified that semantic web technologies and rule-based systems were the most frequently discussed topic within first degree program knowledge management courses. Semantic web technologies and data mining approaches were the most frequently discussed topic within master's degree program knowledge management courses. Secondly, we detected a strong influence of technical-orientation for higher education institutions for the syllabus of knowledge management-related courses. These courses mostly dealt with topics covered by the engineering perspective of the knowledge management domain. Topics covering the organizational and business perspective of knowledge management, such as collaboration, and intellectual property management, are addressed relatively peripherally.

In practice, the research findings of the research can be practically applied in the process of introducing new knowledge management-related courses in a study program or in the process of reengineering existing courses at higher education institutions. From our research it is evident which topics of the area of knowledge management are within courses particularly exposed and for which knowledge management-related topics would be necessary to consider whether they need additional attention in the study process.

After the identification of "hot topics" of knowledge management taught at universities, it would be reasonable to continue the study. In the continuation of the study the issue of inter-rater reliability should be addressed. The inclusion of more rates would decrease the possibility of errors due to the subjective interpretation of the analyzed syllabuses. For the evidence of coverage of the knowledge management spectrum, it appears that all domains of the knowledge management spectrum do not get excessive attention within the study process. For that reason it would be reasonable to verify if the identified "hot topics" of knowledge management domain actually support all the required skills and competences expected by businesses and if the inclusion of additional topics in study processes at universities is required. This question leave us with some research challenges for future studies.

References

1. Al-Hawamdeh, S.: Designing an interdisciplinary graduate program in knowledge management. J. Am. Soc. Inform. Sci. Technol. **56**, 1200–1206 (2005)
2. Argamon, S., Goharian, N., Grossman, D., Frider, O., Raju, N.: A specialization in information and knowledge management systems for the undergraduate computer science curriculum. In: International Conference on Information Technology: Coding and Computing, vol. 1, pp. 476–481 (2005)
3. Bedford, D.: A case study in knowledge management education – historical challenges and future opportunities. Electron. J. knowl Manage. **11**(3), 199–213 (2013)
4. Binney, D.: The knowledge management spectrum – understanding the KM landscape. J. Knowl. Manage. **5**, 33–42 (2001)
5. Dalkir, K.: Knowledge Management in Theory and Practice, pp. 1–24. Elsevier Butterworth-Heinemann, Oxford (2005)
6. Earl, M.: Knowledge management strategies: toward a taxonomy. J. Manage. Inf. Syst. **18** (1), 215–233 (2001)
7. Fang, R., Lin, C., Chang, Y., Tsai, H.: Integrating KM learning Activities into Business Management Curriculum Instruction. MUSP 2008 In: Proceedings of the 8th WSEAS International Conference on Multimedia systems and signal processing, pp. 164–169 (2008)
8. Jennex, M.E.: Classifying knowledge management systems based on context content. In: Proceedings of the 39th Hawaii International Conference on System Sciences, pp. 1–8 (2006)
9. Jing, F., Chakpitak, N., Goldsmith, P.: A knowledge supply chain: reengineering e-tourism curriculum design. E-rev. Tourism Res. **10**(2), 56–62 (2012)
10. Jundale, S., Navale, G. S.: Knowledge management in education. In: Proceedings of the International Conference on Intelligent Agent & Multi-Agent Systems, pp. 1–3 (2009)
11. Kaiser, J. M., Conrad, J., Koehler, C., Wanke, W., Weber, C.: Classification of Tools and Methods for Knowledge Management in Production Development. Design 2008. In: proceedings of the 10th International Design Conference, pp. 809–816 (2008)
12. Kakabadse, N.K., Kakabadse, A., Kouzmin, A.: Reviewing the knowledge management literature: towards a taxonomy. J. Knowl. Manage. **7**(4), 75–91 (2003)
13. Liyanage, S., Poon, P.S.: Technology and innovation management learning in the knowledge economy. J. Manage. Dev. **22**(7), 579–602 (2003)
14. Macintosh, A., Filby, I., Kingston, J.: Knowledge management techniques: teaching and dissemination concepts. Int. J. Hum Comput Stud. **51**, 549–566 (1999)
15. Petrides, L.A., Nodine, T.R.: Knowledge Management in Education: Defining the Landscape, pp. 1–30. Institute for the Study of Knowledge Management in Education, Half Moon Bay (2003)
16. Wilcox, L.C.: The role of expert systems in integrated curriculum design. Expert Syst. Appl. **11**(1), 1–11 (1996)
17. Wiig, K.M.: knowledge management: where did it come from and where will it go? Expert Syst. Appl. **13**(1), 1–14 (1997)

Author Index