

Serap Kurbanoglu Umut Al
Phyllis Lepon Erdoğan Yaşar Tonta
Nazan Uçak (Eds.)

Communications in Computer and Information Science

96

Technological Convergence and Social Networks in Information Management

Second International Symposium
on Information Management in a Changing World, IMCW 2010
Ankara, Turkey, September 2010
Proceedings

Serap Kurbanoglu Umut Al
Phyllis Lepon Erdođan Yařar Tonta
Nazan Uçak (Eds.)

Technological Convergence and Social Networks in Information Management

Second International Symposium
on Information Management in a Changing World
IMCW 2010
Ankara, Turkey, September 22-24, 2010
Proceedings

Volume Editors

Serap Kurbanoğlu
Hacettepe University
Department of Information Management
Beytepe, Ankara, Turkey
E-mail: serap@hacettepe.edu.tr

Umut Al
Hacettepe University
Department of Information Management
Beytepe, Ankara, Turkey
E-mail: umutal@hacettepe.edu.tr

Phyllis Lepon Erdoğan
Bilkent University
Office of the Board of Trustees
Bilkent, Ankara, Turkey
E-mail: phyllis@bilkent.edu.tr

Yaşar Tonta
Hacettepe University
Department of Information Management
Beytepe, Ankara, Turkey
E-mail: tonta@hacettepe.edu.tr

Nazan Uçak
Hacettepe University
Department of Information Management
Beytepe, Ankara, Turkey
E-mail: ucak@hacettepe.edu.tr

Library of Congress Control Number: 2010934653

CR Subject Classification (1998): H.3, H.4, H.5, J.1, H.2.8, K.4

ISSN 1865-0929
ISBN-10 3-642-16031-X Springer Berlin Heidelberg New York
ISBN-13 978-3-642-16031-8 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

springer.com

© Springer-Verlag Berlin Heidelberg 2010
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper 06/3180 5 4 3 2 1 0

Preface

“Convergence” is defined as the intertwining of species or technologies. “Technological convergence,” on the other hand, refers to a trend where a single product such as a cell phone, used in the past solely for communication, evolves into a product that functions not only as a communication device but incorporates the distinct functionalities of a number of other technologies, thereby enabling users to take pictures, listen to music, access the Web, send and receive e-mail messages, find their way, and so on, equally successfully.

Social networks such as Facebook, YouTube, MySpace and LinkedIn, where users congregate, discuss certain issues, entertain themselves, and share information in textual, audio and video formats, are among the most frequented web sites. Social networks having Web 2.0 features offer personalized services, allowing users to incorporate their own content easily and describe, organize and share it with others, thereby enriching users’ experience. More often than not, a capable cell phone is all you need to get access to such social networks and carry out all those tasks. Such tools tend to change our private, social and professional lives and blur the boundaries among them. In other words, our private, social and professional lives are converging, too: someone using a cell phone could be communicating with his/her friend(s), accessing information services, taking an exam using a learning management system, or conducting business.

Needless to say, technological convergence and social networks are also transforming scholarly electronic publishers, library and information centers, and institutional archives. They are no longer “brick and mortar” businesses only, as they provide access to information resources and services on a 24/7 basis to users not even coming to the library building, thereby removing the temporal and spatial barriers. However, users demand more: they expect libraries to offer constant connectivity, communications and content, and to be as accessible, flexible, open to collaboration and sharing as social networks. They expect them to offer more synthesized, specialized and mobilized services at the point of need, without forcing them to change their work, study or social environment. This is a daunting task for information managers, archivists, museum curators and publishers.

The 2nd International Symposium on Information Management in a Changing World, organized by the Department of Information Management of Hacettepe University, took place in Ankara, Turkey, from September 22–24, 2010. The theme of the symposium was “The Impact of Technological Convergence and Social Networks on Information Management.” More than 40 papers were submitted. All papers were subjected to a double-blind reviewing process and 20 were selected for inclusion in this proceedings book. Accepted papers come from 12 different countries and address a number of issues dealing with, among others, digital rights challenges, information literacy, organization of learning spaces, competency of academic library staff, users of digital libraries, data mining of Facebook users’ profiles, digital culture, digital

socialism, personal information management, and automatic categorization, all in the context of information management in the digital age.

We would like to take this opportunity to thank the symposium keynote speaker, Dr. Joan K. Lippincott of the Coalition for Networked Information (USA), and the members of the international Organizing and Program Committees and the Local Committee who invested their time generously to make this event happen. We are most grateful to Phyllis Lapon Erdoğan for editing the final manuscript. We also thank our colleagues Orçun Madran and Erol Olcay who designed the symposium web site as well as the art work; İrem Soydal, Güleda Düzyol, Zehra Taşkın and Tolga Çakmak for carefully copy-editing the papers. Last but not least, it is a pleasure to thank our sponsors and supporters, whose names and logos are listed on the preliminary pages of this proceedings book.

July 2010

Yaşar Tonta
Serap Kurbanoglu

Organization

The 2nd International Symposium on Information Management in a Changing World, was organized by the Department of Information Management of Hacettepe University.

Organizing Committee

Yaşar Tonta	Hacettepe University, Turkey (Chair)
Gülbin Baydur	Hacettepe University, Turkey
İrfan Çakın	Hacettepe University, Turkey
Milena Dobрева	Strathclyde University, UK
Phyllis Lapon Erdoğan	Bilkent University, Turkey
Serap Kurbanoğlu	Hacettepe University, Turkey
Mehmet Emin Küçük	Aksaray University, Turkey
Özgür Külcü	Hacettepe University, Turkey
Micheal Mac an Airchinnigh	Dublin University, Ireland
Orçun Madran	Başkent University, Turkey
Erol Olcay	Turkey
İnci Önal	Hacettepe University, Turkey
John Regazzi	Long Island University, USA
Jordan M. Scepanski	North Carolina University, Chapel Hill, USA
Nazan Özenç Uçak	Hacettepe University, Turkey
Bülent Yılmaz	Hacettepe University, Turkey

Program Committee

Serap Kurbanoğlu	Hacettepe University, Turkey (Chair)
Mustafa Akgül	Bilkent University, Turkey
Buket Akkoyunlu	Hacettepe University, Turkey
Umut Al	Hacettepe University, Turkey
Arif Altun	Hacettepe University, Turkey
Nazlı Alkan	Ankara University, Turkey
Bekir Kemal Ataman	Marmara University, Turkey
Doğan Atılğan	Ankara University, Turkey
Ágnes Hajdu Barát	University of Szeged, Hungary
Carla Basili	Ceris Institute, Italian National Research Council, Italy
Gülbin Baydur	Hacettepe University, Turkey
Yıldıran Bitirim	Eastern Mediterranean University, TRNC
Albert Boekhorst	Amsterdam University, The Netherlands
Michael Buckland	University of California, Berkeley, USA
Mehmet Canatar	İstanbul University, Turkey

Banu Cangöz	Hacettepe University, Turkey
Leslie Chan	University of Toronto, Canada
Kürşat Çağiltay	METU, Turkey
İrfan Çakın	Hacettepe University, Turkey
Ahmet Çelik	Hacettepe University, Turkey
Milena Dobрева	Strathclyde University, UK
Aydın Erar	Mimar Sinan Fine Arts University, Turkey
Levent Ertürk	Atılım University, Turkey
Maria Francisca Abad Garcia	Valencia University, Spain
Chris Hagar	Dominican University, USA
Suliman Hawamdeh	Oklahoma University, USA
Forest Woody Horton	International Library and Information Consultant, USA
Aleksandra Horvat	Zagreb University, Croatia
Maija-Leena Huotari	University of Oulu, Finland
Nikola Ikonov	Bulgarian Academy of Sciences, Bulgaria
Oğuz İcimsoy	Marmara University, Turkey
Ian M. Johnson	The Robert Gordon University, UK
Leif Kajberg	Independent Researcher/Adviser, Denmark
Hamza Kandur	Marmara University, Turkey
Rajkumar Kannan	Bishop Heber College, India
Sekine Karakaş	Ankara University, Turkey
Asker Kartarı	Hacettepe University, Turkey
Hasan Keseroğlu	Istanbul University, Turkey
Adolf Knoll	The National Library, Czech Republic
Dincho I. Krastev	Central Library of the Bulgarian Academy of Sciences, Bulgaria
Mehmet Emin Küçük	Aksaray University, Turkey
Özgür Külçü	Hacettepe University, Turkey
Jesús Lau	USBI VER Library Veracruzana University, Mexico
Aira Lepik	Tallinn University, Estonia
Micheal Mac an Airchinnigh	Dublin University, Ireland
Jeppe Nicolaisen	Royal School of Library and Information Science, Denmark
İnci Önal	Hacettepe University, Turkey
Fahrettin Özdemirci	Ankara University, Turkey
John Regazzi	Long Island University, USA
Angela Repanovici	Transilvania University of Braso (Romania)
Fernanda Ribeiro	Porto University, Portugal
Jordan M. Scepaniski	Jordan Wells Associates, USA
İrem Soydal	Hacettepe University, Turkey
Fatoş Subaşıoğlu	Ankara University, Turkey
Yaşar Tonta	Hacettepe University, Turkey
Nilüfer Tuncer	Hacettepe University, Turkey
Nazan Özenç Uçak	Hacettepe University, Turkey
Yurdagül Ünal	Hacettepe University, Turkey

Sponsors



Hacettepe University



Anatolian University Libraries
Consortium



EBSCO Information Services



Embassy of the United States of
America Ankara, Turkey



Emerald



Mikro Bilgi Kayıt ve Dağıtım
A.Ş.



PECYA



Springer



Turkish Librarians' Association



Association of University and
Research Librarians

Table of Contents

Technological Convergence and Social Networks

“My Information:” Digital Libraries, Social Networking, and the User Experience (Keynote Paper).....	1
<i>Joan K. Lippincott</i>	
The Implications of Information Democracy and Digital Socialism for Public Libraries.....	3
<i>Esin Sultan Oğuz and Leif Kajberg</i>	
Collaboration through Communities of Practice in the Digital Age.....	18
<i>Fatih Oguz, Corrie V. Marsh, and Cliff Landis</i>	
Let Me Tell You Something about (Y)our Culture?.....	31
<i>Mícheál Mac an Airchinnigh</i>	
Digital Natives and Specialised Digital Libraries: A Study of European Users.....	45
<i>Milena Dobрева, Emma McCulloch, Duncan Birrell, Yurdagül Ünal, and Pierluigi Feliciati</i>	

Academic Libraries and Information Literacy in the Digital Age

Impact of Recent Trends in Information and Communication Technology on the Validity of the Construct Information Literacy in Higher Education.....	61
<i>A.A.J. (Jos) van Helvoort</i>	
Exploring the Competency of Academic Library Staff Engaging in Emerging Services.....	74
<i>Haruki Nagata</i>	
Programming New Learning Spaces: The Changing Nature of Academic Library Buildings.....	85
<i>H. Lea Wells and Jordan M. Scepaniski</i>	
The Role of Information Literacy in Environmental Scanning as a Strategic Information System - A Study of Singapore SMEs.....	95
<i>Xue Zhang, Shaheen Majid, and Schubert Foo</i>	

Information Management and Digital Rights Challenges

Using Internet Services for Personal Information Management	110
<i>Shaheen Majid, Myint Malar San, Saw Tin Naung Tun, and Thu Zar</i>	
New Approach for Automated Categorizing and Finding Similarities in Online Persian News	120
<i>Naser Ezzati Jivan, Mahlagha Fazeli, and Khadije Sadat Yousefi</i>	
An Evaluation of TS13298 in the Scope of MoReq2	129
<i>Gülten Alır, Thomas Sødning, and İrem Soydal</i>	
Identification of User Patterns in Social Networks by Data Mining Techniques: Facebook Case	145
<i>A. Selman Bozkir, S. Güzin Mazman, and Ebru Akçapınar Sezer</i>	
Digital Rights Challenges in a World of Technological Convergence	154
<i>John N. Gathegi</i>	
Copyright Issues Related to the Digitization of Cultural Heritage in Croatia	161
<i>Aleksandra Horvat and Daniela Zivkovic</i>	

Information Systems and Bibliometrics

Performance Evaluation of Web Information Systems	175
<i>Yaşar Tonta and İrem Soydal</i>	
Information Science and Cognitive Psychology: A Theoretical Approach	189
<i>Armando Malheiro da Silva, Fernanda Ribeiro, and Fernanda Martins</i>	
Visualizing the Structure of Scientific Output of Iranian Scholars in Science Citation Index (SCI) during 2000-2006	200
<i>Farideh Osareh and Maryam Keshvari</i>	
A Study on Scientific Product of the University of Tehran in Web of Science Database during 1989 – 2009	211
<i>Farideh Osareh and Amin Zare</i>	
Contribution of Turkish Scholars to Earthquake Literature: The Impact of the Marmara Earthquake	222
<i>Zehra Taşkın</i>	
Author Index	231

“My Information:” Digital Libraries, Social Networking, and the User Experience

(Keynote Paper)

Joan K. Lippincott

Coalition for Networked Information, 21 Dupont Circle, Washington, DC 20036
joan@cni.org

Extended Abstract

The constituents of our universities, whether they are students or members of the academic faculty, are using information systems in a wide variety of ways. Information is woven into their lives through the myriad systems that they access and the social networks in which they participate. At present, most individuals who work in the academic environment own a variety of devices and use them for varying purposes; most of their devices are used for more than one activity. For example, a mobile phone may be used to telephone others, to send and receive text messages, to take photos, and to wake up the owner in the morning with the alarm clock capabilities. It is uncertain at this stage of development whether individuals will continue to own and use a variety of devices or whether they will prefer to adopt one or two devices which allow them the functionality they need in a convenient physical package.

Increasingly, academics and students will use their devices to access, store, and create information, in text, still and moving images, sound, and geo-spatially referenced digital objects. In effect, they will create and enhance personal information environments. Using mobile and desktop devices that either store or link to their favorite information resources or provide easy access to new ones, they will collect information, communicate, create new products, and share them with colleagues, family, and friends.

Will libraries, archives, and museums be integral resources in these environments or will they be peripheral players? Understanding user needs and behaviors will be key to developing information environments and services that enhance academic work and delight the imaginations of a wide variety of users. For example, for someone using a mobile device, what does it mean to “find a book?” Might that user want to look up the book in a catalog, find its location, query whether it is already checked out, and get directions to locate it within a physical library? Or, might that user want to search a catalog or database, authenticate to the institution’s authorization management system (assuming it is a licensed resource), download it in a format compatible with his or her mobile device, and find out what respected colleagues and friends think about that resource? Information professionals need to question assumptions about users’ needs and behaviors as they design systems for users of mobile devices.

The integration of various types of functionality will be a challenge for libraries and museums. If our users want mechanisms to customize their own means of access to library resources, linking those resources to citation software, to course management systems, to Facebook, and other means of connection to others, information professionals must ensure that they develop systems with interoperability and portability of information.

Users will expect customized and customizable information environments, and information professionals need to have as clear as possible understandings of their needs. For example, information professionals in their institutions should use a variety of methods, e.g. user surveys, log analysis, web analytics, focus groups, and interviews to find information about:

- Ownership of and access to types of devices
- Types of activities they engage in on their devices
- Ways they would like to connect disparate collections of information
- Types of interfaces they prefer
- Types of social networking tools they regularly use
- Types of academic or course management tools they regularly use
- Cloud services with which they connect
- Means of accomplishing collaborative work

Libraries and museums are developing some creative services to better address the needs and style of today's technology-savvy users and visitors. For example, museums may employ QR codes to allow visitors to find more information about art objects; libraries may use tag clouds as a mechanism by which a user can refine a search; catalogs may include links by which a user can get a reference in a specified style for a bibliography. Museums and libraries are encouraging users to enrich metadata and to add their personal stories to photos in FlickrCommons. Libraries and museums are developing geo-referenced collections that enable users to walk around a city or other area and link to library or museum information about specific places of interest.

Many libraries and museums develop applications and services that reflect the way that they see their organizational operations, e.g. looking something up in a library catalog or visiting a museum. In the next phase of the information revolution, we will see more emphasis on integrating the content and services of information providers into users' personal information environments. At present, many of the applications that help make our users more efficient in their use of information or assist them in developing new approaches to their questions and problems operate somewhat in isolation. The challenge for libraries and museums is to develop applications related to their collections and services that easily integrate into users' preferred personal information environments.

The Implications of Information Democracy and Digital Socialism for Public Libraries

Esin Sultan Oğuz¹ and Leif Kajberg²

¹ Hacettepe University Department of Information Management,
06800 Ankara, Turkey
esinsultan@hacettepe.edu.tr

² Independent Researcher/Adviser, DK-2880 Bagsvaerd Copenhagen, Denmark
leif.kajberg@gmail.com

Abstract. In these times, public libraries in many countries have increasingly come under pressure from developments within the information landscape. Thus, not least because of the massive digitization of information resources, the proliferation and popularity of search engines, in particular Google, and the booming technologies of Web 2.0, public libraries find themselves in a very complex situation. In fact, the easy-to-use technologies of Web 2.0 challenge the basic principles of information services provision undertaken by libraries. The new digital information environment and social software tools such as blogs, wikis and social networking sites have fuelled a discussion of the future of public libraries as information providers. After all there seems to be a need for public libraries to reorient their aims and objectives and to redefine their service identity. At the same time search engines, and especially Google, are increasingly coming under scrutiny. Thus, analysis results referred to show that the conception of information and the underlying purpose of Google differ from those of public libraries. Further, an increasing amount of criticism is being directed at collaborative spaces (typically Wikipedia) and social networks (e.g. MySpace) and it is pointed out that these social media are not that innocent and unproblematic. In discussing the survival of public libraries and devising an updated role for libraries in the age of Google and social media, attention should be given to fleshing out a new vision for the public library as a provider of alternative information and as an institution supporting information democracy.

Keywords: Collective intelligence, information democracy, public libraries, Web 2.0.

1 Introduction

Digitization of information combined with increasing growth of electronic networks has created new opportunities for providing information resources and services for citizens. New forms of and channels for distributing information and documents within the Internet, new tools and opportunities for digitizing our written cultural heritage and making it accessible, new mechanisms for discovering and accessing

information and new services and networking forums such as Facebook, Flickr, MySpace, Twitter, YouTube and social tagging provide an opportunity and challenge to the kind of services the public library offers and to the societal role and institutional identity it assumes for itself. Activities and services facilitated by the Internet are increasingly used by various citizen groups. Especially younger generations have embraced the new forms of electronic interaction and adopted Internet as their own media. Thus, harnessing the benefits and challenges of Web 2.0 remains a major challenge to public libraries today.

The term “2.0” is just a shorthand used to label an unknown but desired move towards something new. The change is evolutionary and relatively gradual in a world where it is almost a full-time job to stay up to date with new daily developments. There is no single and unambiguous definition of Web 2.0, although certain experts are capable of explaining its characteristics. Tim O’Reilly, founder of the leading publishing house for computer and networking publications, suggested these features as some of the key elements of a Web 2.0 application: it takes place on the Web; it is a service, not a product; it is not limited to a single software product or a single machine; it is open and shared; users in groups along with social interaction are part of its organization; users contribute content and add value [1]. But the 2.0 concept is not just about *searching*. The last item or aspect included in Tim O’Reilly’s definition of Web 2.0 above is that users participate and add content and value. The first decade of the present millennium will probably be known for the expansion of sophisticated digitally-based social activity. Users have become accustomed to creating content on the Web, be it drafting a review of a book at an online bookstore or creating an identity for themselves on MySpace. They are also familiar with articulating themselves by posting comments to blogs or contributing their ideas on a topic to Wikipedia. Today’s web users have an expectation that they will find a community at their chosen electronic destination. They also expect to interact with their information resources, not to consume them in a passive mode [1].

1.1 Methodology

The study presented here aims to analyse the ways in which public libraries can strengthen their survival capacity by drawing upon the new Web 2.0 technologies available and develop new roles. An analysis is conducted of selected writings covering such key notions as social software applications, collective intelligence and digital socialism. Also revisited is the dated concept of *the political library*. Based on observations emerging from the analysis, a revised role is outlined for public libraries in the era of digital information and Web 2.0 with a special focus on information democracy and the library’s function as a neutral information provider in a Google-dominated commoditized information world. Besides, the intention is to shed light on the nature, viability and conditions and opportunities of information democracy within the framework of today’s social networking media.

1.2 Literature Review

According to Stark [2] the ability of users to produce and disseminate new creations and take part in public cultural discourse is called semiotic democracy which is

simply named Web 2.0 or social web. In the current information age, the capacity of a society to effectively position itself as a consumer and producer of knowledge is crucial to its social and economic development. Today, the co-relation between knowledge and development appears to be well established. Increasingly frequently, social, economic, and political progress are linked with the ability of countries to make informed decisions and knowledge-based choices [3].

The perception of information technology as fundamental development tools of the 21st Century rests on several assumptions. At the macro level, it assumes that the introduction and use of technologies will improve the efficiency of developing countries' industrial infrastructure, enhance their overall economic performance, and strengthen their competitive capacities in the global market [4]. At the human development front, it assumes that information technologies will contribute to improvements in the provision of basic social services, help disseminate valuable information on production and conservation, improve the efficiency of governments, and enhance the provision of education and health services [5]. In other words, information and communication technologies are promoted as windows through which greater access to outside expertise and almost unlimited access to a wealth of knowledge and information - which otherwise will remain inaccessible - will be obtained. Some go even further to argue that the introduction of information technologies will contribute to the strengthening of democracy; increasing social participation; and removing barriers to modernization, making populations fuller agents in their development and members of the "global village" [6].

The theoretical framework provided for the present study draws on inspiration from Doctor's piece on justice and social equity in cyberspace [7]. This article was published in the early days of the Internet characterised as they were by enthusiasm, euphoria and a fascination of the promising new potentials and possibilities represented by the new global medium and utility. Revisiting the somewhat idealistic ideas, notions, conceptions and projections that arise in the first, pioneering and booming years of the Internet from a contemporary information democracy perspective is one of the objectives of the analysis reported here. Characteristic to the pioneering years of the Internet and "the Information Super Highway" is the fairly optimistic and in some respects even idealistic tone. Today, things are more complex and we are seeing the commercialisation of the Internet, "abuse" of the Internet (hacking, theft of money and identities and other types of crime) along with the "hedonistic" take-over (e.g. through the spread of porn) so effectively depicted by Keen [8].

Current professional literature on the implications of Web 2.0 technologies for libraries and their service provision tend to emphasize the new social software tools and media as information assets to be integrated into existing service offerings. The Web 2.0 social media are typically seen as opportunities and means for supplementing, enhancing and enriching the existing mix of library-related services and facilities. Briefly, Google and interactive technologies such as wikis and blogs are considered new devices in the library service provision toolbox. Eisenberg provides an overview of what he identifies as representative cutting-edge/web 2.0 technologies and singles out 3-D virtual worlds (Second Life), social networks (MySpace and Facebook), personal digital devices (iPod, PDA, BlackBerry, cell phone) for special treatment. A

SWOT analysis is conducted of the options selected and in pointing to their opportunities, Eisenberg [9] explains how libraries can make the most of these innovative web-based tools in providing services and enhancing access to resources. Quite a few authors of articles stress the proactive position and the generally positive attitude to be adopted by libraries in availing themselves of the new opportunities and facing the challenges provided by Web 2.0 technologies and user patterns. Lietzau [10], for instance, reports a study undertaken by Colorado State Library's Library Research Service on the pace at which Web 2.0 tools are finding their way into American public libraries. The statistical information included also gives an idea about how the various social web applications are used in libraries. It appears that chat reference is the most prevalent service. Not surprisingly, larger library systems were prime movers in this area and more likely to appreciate and make use of the new Web 2.0 opportunities than smaller ones. Lancaster [11] discusses how libraries can profit from Web 2.0 by integrating social networking tools in the library's online presence. For instance, the new web-based social media should be seized as an obvious opportunity for enhancing the library's visibility and community profile and for reaching out to new user groups. Based on the results of a qualitative study Luyt, Ally, Low and Ismail [12] discuss the relevance, usefulness and weaknesses of Wikipedia as seen from a practicing librarian's perspective. The aim of the study was to identify the perceptions of a group of Singapore librarians with respect to Wikipedia as an information source to be consulted in reference services and the actual use made of this tool in servicing users. One finding is that the librarians participating in the study have a balanced view of Wikipedia and they are well informed about the limitations and shortcomings of this collective space product. The librarians interviewed also know about the negative views of Wikipedia and the ongoing critical discussion of the tool, but they value its coverage of non-Western topics. Thus, Wikipedia is evaluated positively for the information it contributes on more Asian-centric, multi-cultural and multilingual issues and aspects.

However, there are signs that a more critical awareness of Web 2.0 phenomena is beginning to gain ground. Brabazon [13] has some serious reservations about the whole ideology behind and the peer production practices of Wikipedia and she is very concerned about what Google does to students in pursuing projects and assignments. Based on recorded examples from classes, she shows how reliance on Google impacts negatively on students' informational behavior and makes it increasingly difficult to cultivate a habit of sound scholarship. In a very thoughtful piece, Waller [14] takes a close look at the relations between Google and public libraries and explores similarities and differences. On the surface of it, Google seems to pursue goals and offer services and products that are parallel to or overlap the kinds of searching assistance and information provision that are core activities in libraries, but in the end the two players in the information arena deviate markedly from each other. The author demonstrates that the *conceptions of information* adhered to by (1) Google as a commercial firm and (2) public libraries as providers of balanced and consolidated information are fundamentally different. The commercial firm and the public agency simply want to do different things. It is argued that public libraries are concerned with the content of information and that libraries have as their purpose to provide access to

information with the wider policy of strengthening information democracy. In contrast, Google considers information a vehicle for advertising, and information provision and refinement of search capabilities in a Google perspective boil down to facilitating targeted advertising. In this respect mention is made of Google's systematic efforts to gather all sorts of information about users by tracking consumption habits. A big brother is watching you-like scenario does not seem far away. Waller's reflections on the democracy-underpinning role of public libraries in maintaining a balanced and non-commercial information provision are very central to the observations on a redefined role for public libraries in the present paper.

A decidedly pessimistic view of Web 2.0 and interactive social media can be found in Keen's book *The Cult of the Amateur* [8]. The book embodies a frontal attack on what the author sees as the frightening regime of amateurs and a pervasive culture of narcissism resulting from the Web 2.0 revolution. Keen provides a coherent and very critical perspective on the web 2.0 tools and phenomena and demonstrates their manipulating potentials and how they erode expert knowledge and expert performance and gradually bring about de-professionalization in some respects. Professionals have been replaced by noble amateurs. Keen explores the seamy side of blogs and blogging and addresses the problem of tricksters and fraudulent behavior. He provides examples of dubious editorial practices characterizing Wikipedia and the mediocrity of content provided by contributors. Above all he laments the downgrading and dismissal of experts and the devaluation of expert knowledge. Keen ends up with a very pessimistic state-of-the-art description and scenario in which he identifies Orwell-like tendencies and points out that the American society is moving into an age of total digital surveillance. Sounding a bit like an old moralizing culture critic, Keen draws attention to a range of critical and pertinent issues affecting all web users.

The published literature is very sparse on the implications of Web 2.0 and social networking for the community involvement of public libraries. Actually, very few contributions address the role of libraries in maintaining freedom of information in the Digital Age along with their supportive role in relation to campaigning initiatives, local grassroots activities, the organisation of political debates as well as the provision of alternative, anti-mainstream and anti-elitist information, etc.

In contrast, library literature, especially that part of it which covers 20th century developments in libraries and librarianship in Australia, UK and USA, provides considerable coverage of the role of libraries in promoting and consolidating democracy. For instance, Waller [14] refers to what she calls the "grand tradition" of public libraries in the 1950s with Lionel McColvin, UK as one of the leading figures. According to McColvin public libraries would have a leading role in advancing democracy, in knowledge building and the spread of knowledge and in empowering citizens through the possibility for self-education. However, recent library literature also includes items that focus on libraries and democracy and the societal role of libraries. In his monograph on *Civic Librarianship* McCabe explores the concept of civic librarianship and develops a vision for the mission and purpose of the public library. Civic librarianship differs markedly from the libertarian public library, but it is also very different from the public library of the traditional type, which has often fallen short in fleshing out its basic mission into effectual and tangible strategies for action.

McCabe [15] sees a broadened role for public libraries and identifies a number of areas where strategic action is needed:

- Restore the confidence of public librarians and trustees in exercising social authority.
- Renew the public library's historical mission of education for a democratic society.
- Develop the public library as a center of the community.
- Develop strategies to build communities through public library service.
- Use services and collections to meet social as well as individual needs.
- Strengthen the political efforts of public librarians and trustees.

As can be seen, the suggestions for reforming public libraries in line with the conceptual framework of civic librarianship are of a more general nature and since the book appeared in 2001 there is no treatment of the challenges of e.g. social networking technologies and the way people communicate and organize information-related activities *outside* the library context after the advent of the social web revolution. However, it is emphasized that the public library should promote community identity, community dialogue, community collaboration and community evaluation. McCabe urges libraries to adopt an approach and an attitude in line with *communitarianism*, a political movement that emphasizes community, tradition and social order as opposed to extreme individualism. Thus, the library should develop into a centre of community life and promote identity-building, public dialogue and cooperation.

Clearly, civic librarianship is meant as an effort to update and expand the role of the public library while keeping the library's historic mission of education for a democratic society. The author's insistence on civic dialogue and social interaction is also of relevance when discussing and defining the role of the public library in times of web 2.0.

Kranich [16] explains how libraries help reduce the digital divide, increase access to government information and are fighting against both censorship and private interests to ensure that access to information is as free as possible. The library as civic space creates opportunities for community and dialogue, which she thinks is a very important democratic function as a supplement to information-related and education-centred tasks. In their joint article Canadian library researchers Alstad and Curry [17] describe how squares and other public spaces are increasingly replaced by company-owned areas such as shopping malls, where people can no longer act as citizens, but are primarily consumers. In order for libraries to support democracy and serve as public space they should, among other things, change their objectives so that they move towards a more proactive stance thus making room for lectures and discussion groups. A Danish perspective is provided by Skot-Hansen and Andersson [18] who carried out a study of libraries as a resource in the local community. As pointed out in the study, for a library to serve as a local driver it should relate actively to the community it belongs to and sharpen its profile in interaction with other institutions, associations and groups. The libraries' social function is also examined in a British study conducted by Matarasso [19]. He concludes that libraries have a great potential to contribute to the development of the local community. In a contribution in the anthology titled *Libraries and Democracy: the Corner-stones of Liberty* Durrance and

others [20] explore several American library projects that address web-based community information, which are considered to help strengthen civil society. The libraries' own websites can be used successfully, for example, in providing guidance to citizens in pointing to web-based government information resources and be targeted to various minority groups. Also, American libraries have often been leaders in or active as partners in the development of virtual local area networks, so-called community networks. Lately, findings of a large-scale study of those who use public computers and Internet access in public libraries show that the use of library technology had significant impact in four critical areas: employment, education, health and making community connections. According to the results of the study titled *Opportunity for All: How the American Public Benefits from Internet Access at U.S. Libraries* 40% of library computer users (an estimated 30 million people) received help with career needs. Among these users, 75% reported they searched a job online. Half of these users filled out an online application or submitted an online resume. Library computers also have a major impact on linking patrons to their government, communities, and civic organizations [21].

2 Collective Intelligence

People have used the phrase "collective intelligence" for decades and it has become increasingly popular and more important with the advent of new communications technologies. Although the expression brings to mind ideas of group consciousness or supernatural phenomena, when technologists use this phrase they usually mean combining behavior, preferences, or ideas of groups of people to create novel insights. Collective intelligence was of course possible before the Internet. The web is not a required means when gathering data from disparate groups of people, combining it and analyzing it. One of the most basic means of this information gathering activity is a survey or census. Collecting answers from large groups of people lets you draw statistical conclusions about the group that individual members would not have known by themselves. Building new conclusions from independent contributors is really what collective intelligence is about [22].

A book by Pierre Lévy [23] about the computerization of society from a social-theoretical standpoint represents an early approach to *Collective Intelligence*. The author develops, as we see it, a very exciting conceptual framework. He believes that it is feasible to devise and build up sophisticated systems of networked intelligence defined as collective intelligence. These systems include the potential for guiding humanity into a new era of intellectual and social achievements. Collective intelligence, in Lévy's sense, is "a global project whose ethical and aesthetic dimensions are as important as its technological and organizational aspects" [23]. Lévy envisages the creation of a new *knowledge* space that grows out of computer technologies such as hypertext and he refers to this new knowledge space as the *cosmopedia*. The author makes the point that through collective intelligence, societies can continue the project of emancipation initiated during the Enlightenment. However, Lévy's work was written in the early years of the Internet -it appeared in 1997- and parts of his vision for a genuine democratic cyberspace pervaded by human and social ideals appear somewhat dated taking into account how the world looks today. At the "macro level", for

instance, he was not able to predict such phenomena as the climate collapse, the emergence of globally oriented terrorism and the war against it and neo-nationalist tendencies. Further, he tends to underrate the upsurge of entertainment and infotainment, the dominance of commercialism and the spread of hedonistic behavioural patterns. Moreover, technological and social developments have moved so fast during subsequent years that part of the monograph's ideas and messages seem questionable today. On the whole, Lévy's work is very "philosophical" in scope and even metaphysical and it displays obvious features of techno-utopianism. On the other hand, many of Lévy's original and pioneering concepts -including real-time democracies- trigger further thought and invite critical scrutiny in the light of today's notions of collective intelligence. Overall, there is a catching drive in the book and it is very visionary, but there are obscure passages and incomprehensible observations as well. Problematic too are some speculative and prejudiced geo-political statements on the nature of "the South". In the chapter on "the Dynamics of Intelligent Cities" the author explains the idea of a direct, computer-mediated democracy -*a virtual agora*- and he anticipates the application of web 2.0 tools for discussions and decision-making. The author adds that the introduction of what he calls a *real-time mechanism for direct democracy* would facilitate a democratic dialogue. "Within the framework of collective intelligence, real-time democracy is the absolute antithesis of the demagoguery of live action broadcasts and the immediacy of crowd behaviour" [23]. One reservation that might be voiced here is the reaction or response from those representing the establishment and the political scene: are politicians and those in power today really interested in this kind of direct democracy?

3 Digital Socialism

Much more concrete and pertinent is Kelly's [24] intriguing and captivating *Wired* piece on the emergence of what he labels "The New Socialism". New or Digital Socialism refers to the advent of a global collectivist society sustained and nourished by the communication, exchange and knowledge sharing activities undertaken by tens of thousands of voluntary web content producers, maintainers and developers, knowledge exchangers and free agents throughout the world. In their social interactions and in providing their cross border selfless services, they all rely on social technology. Taken together, they all operate within a collective digital culture and represent a socialist workforce. According to Kelly the classic Wikipedia activities, the maintenance of specialized wiki engines, the mushrooming of knowledge sites and file-sharing are examples of an emerging collectivism. Digital socialism has very little to do with old-school or "red-flag socialism" since there is no state, centralized government rigid bureaucracy. As the author puts it, "Instead of faceless politburos, we have faceless meritocracies, where the only thing that matters is getting things done". The new digital mode socialism cannot be categorized as an ideology. "Rather it is a spectrum of attitudes, techniques, and tools that promote collaboration, sharing, aggregation, coordination, adhocracy, and a host of other newly enabled types of social cooperation. It is a design frontier and a particularly fertile space for innovation". Kelly presents clear evidence that collective work modes are in the process of rapid expansion, but a couple of statements echo a slightly over-optimistic tone. It seems

that the author slightly ignores some still existing harsh economic realities including the pervasiveness of neo-liberalism. Socialism in an economic sense not least involves the way economic resources are managed and how income is redistributed in society. But Kelly may be right; perhaps we are moving towards a non-capitalistic, open source and peer-production economy adopting features from both Wikipedia and the moderate socialism of Sweden.

But given that we are seeing the materialization of a society of digital socialism, what will then happen to libraries? Are they just becoming obsolete and marginalized repositories decoupled from the networks and communities of the busy online bees? How can public libraries interact with the new socialists in Kelly's vision? And how can libraries contribute to and qualify the new socialist work modes in a convincing and meaningful way? Could they serve as digital resource centers and support mechanisms in an emerging culture of digital collectivism? There are quite a few problems and issues that need exploration.

4 The Political Library: Revival of a Concept?

In her thesis on the Political Library with the subtitle "Public Library as a space for citizens' participation and public discourse" Jadinge [25] discusses the potential public libraries have for actively supporting civic participation and public discourse. The study seeks to explore the origin of the idea of the political library in a Swedish public library context in the mid-1970s. The author observes that the political library deserves to be taken out of oblivion mainly for two reasons. First, it is an idea that is quite radical (in the general sense of the word!) by today's standards, and it should therefore serve as fuel for a renewed discussion of library ideology and democracy issues, in field practice as well as in research. The concept of a political library is interesting because it affects some fundamental aspects of library and information activities, such as the neutrality/objectivity issue and the relationship that libraries have to civil society. Secondly, it is relevant to offer a historical perspective to today's library debate. The author's view is that undertaking a comparison between the context of the 1970s and the situation and conditions of the 2000s can be fruitful. As is the case today, democracy problems were frequently and sometimes heavily discussed in the 1970s, but the atmosphere and context were different and attention was focused on how the political library should act so as to maintain the library's neutrality. To be neutral may nevertheless often involve some sense of commitment.

First, why should libraries devote themselves to promoting civic participation and public discourse, and secondly, it is at all necessary or desirable to attempt to increase citizen participation in politics?

Based on the conclusions of the study it is believed that libraries have an important role in promoting the local community. Libraries do not consider their users as uninterested in community information and policy. But this given, why do libraries not to a greater extent encourage and stimulate public dialogue, opinion building, advocacy and social dialogue? The investigator's conclusion is that for the moment Swedish public libraries do not really consider civic participation and public conversation a significant priority.

The results of the Swedish study prompt further analysis of the notion of the political library, its relevance today along with its potential for renewing the role of a public library in transition. Today, appraising the generalizability and pertinence of the political library and giving the concept a needed brush-up implies an awareness of the opportunities of web 2.0 tools and applications.

5 Access to Alternative Information

Given that public libraries take their function as provider of alternative, non-elitist and non-mainstream information seriously, there are many situations where the active involvement and service provision of libraries would be relevant and desirable. Illustrative examples are the campaigns and debate sessions preceding elections, referendums, etc. Typically, and this observation could be generalized to many countries, the official information presented to the electorate is biased. Thus, for instance, in Denmark the many referendums relating to Denmark's entry into the Common Market and the EU as well as Denmark's accession to the EU treaties, etc. constitute an illustrative example: there is unequal access to information and lack of funds for distributing alternative information. Frequently, there is a marked lack of alternative information resources reflecting positions other than those held by the establishment and those possessing the political power and the money. There is a need for information that provides alternatives to and challenges the official and dominating messages and viewpoints. The new social network media have partly remedied this situation, but libraries could still play a role here.

5.1 Facebook as an Information Tool for Local Protest Actions: A Danish Example

In Denmark the controversy over and the fight for the survival of a local railway in a thinly populated area provides an illustrative example of the involvement or lack of involvement of the local public library in a much discussed local matter. For the time being the Western railway, a local railway line in the Western part of Denmark, is at risk of being closed down in that a majority of Regional Council Members want to eliminate the line because it is considered economically loss-making; it is argued that it is too expensive in terms of operational and maintenance costs and the case is made that buses are a better solution. The prospect of a rural railway line ceasing to exist because of a Regional Council decision evoked strong protests from parts of the local population, created a heated debate and led to the formation of railway protection initiatives. Also, a group on Facebook named "Save the Western Railway" was set up. However, the local library has adopted a fairly passive role in relation to the railway issue. No meetings have been hosted by the library and the only activity organized by the library is the setting up of an exhibition featuring the railway and its history. The Western Railway protection citizens' initiative represents an interesting case illustrating how Facebook is relied on by politically articulate individuals and groups. There are tens of thousands of examples of this nature on Facebook. These grassroots activities, campaigns, protest groups and unofficial networks confronting decision-makers and those in power provide examples of how initiatives are born, strategies are developed, individuals get involved

and become members of groups, how communication takes place, how various types of information and views are presented and exchanged and how decisions are made, etc. Also illustrated are the exchange of information, views, advice and know-how between various bodies of expertise and those who maintain grassroots initiatives. And last but not least: studies of the emergence of grassroots initiatives in a Facebook context -or as they develop within other social networking media- could be designed so as to explore the ways in which libraries respond to, support or ignore groups and initiatives arising and developing within the social networking media.

There are various ways in which public libraries could adopt a more proactive role in relation to Web 2.0 and citizens' campaigns and initiatives. Thus, a Danish project, outlined on the web pages of the Librarians' Union, addresses the role of the public library as a moderator of current political debates, etc. going on in the local community. The library is supposed to provide balanced subject-specific input for discussions progressing in social network media of the Web 2.0 type. You can have people debating current and crucial topics and issues on the Web. But the prerequisite is that you prepare solid background information and that you dare bring up controversies, tender subjects and sensitive issues for discussion. Also, you should be ready to interact with other media. On the whole, libraries could adopt a more active democratic role.

5.2 YouTube as a Censored Tool for the On-Line Community: A Turkish Example

In Turkey, the popular video-sharing website YouTube has been blocked since 2008 after a series of insults between the Greek and Turkish users of the site escalated. The row between Turkish and Greek YouTube users started when Greek videos claimed that Atatürk, the founder of the Turkish Republic, and the Turkish people were homosexual and showed images of the Turkish flag overlaid with profanities. This dispute received a lot of attention in the Turkish media, which they labeled a *virtual war* with both sides posting insulting videos. YouTube agreed to take down the offending videos, but nevertheless, the prosecutor in Istanbul got a court order, based on charges of insulting Atatürk, which is illegal in Turkey. YouTube responded that it had taken down the videos and was cooperating with the government, adding that "while technology can bring great opportunity and access to information globally, it can also present new and unique cultural challenges". Under article 301 of the Turkish penal code, public denigration of Turkishness, The Republic of Turkey, Atatürk and the other national symbols is punishable by imprisonment. The following message appears on the screen after attempting to visit YouTube: "*The decision no 2008/402 dated 05.05.2008, which is given about this web site(youtube.com) within the context of protection measure, of Ankara 1. Sulh Ceza Mahkemesi has been implemented by Telekomünikasyon İletişim Başkanlığı.*" The largest Internet provider and privatized state company has put the order into effect while some smaller Internet providers still allow access to the site. The head of Turk Telekom has said that he is in no position to judge the nature of the videos posted, but is simply following court orders to enforce the ban. About 50 writers in the country have been put on trial for allegedly contravening the rule, though most cases have eventually been dismissed by the judge. It is also unclear how long the ban will last [26].

5.3 Libraries and Discussion Forums

Thus, as illustration, in the Danish Municipality of Odder it has for several years been natural for citizens and politicians to engage in discussions on a variety of issues using web based discussion forums. Last year's municipal elections provided another example of the electronic communication between citizens and local politicians in that more than 400 comments were posted as part of a lively debate between citizens and those standing as candidates for the town council. One of the reasons for the recorded success in raising and maintaining e-debates is that those responsible for hosting and maintaining the debate invest quite a lot of effort in furnishing people with background knowledge on a specific topic or issue. For instance, all town council decisions are described in a journalistic mode on the commune homepage. In addition, video transmissions of sequences selected from, among other things, town council meetings and local civic meetings on key issues are available. It is crucial to bring up tender subjects and sensitive issues for discussion. If you dare not put something on the line and raise a controversy in areas and issues people are very eager about they tend to drop out and ignore debates.

Unfortunately, most local authorities and councilors tend to avoid conflict and shrink from raising sensitive subjects. Thus, it is obvious that the initiative rests with the libraries when it comes to providing local residents with opportunities for making themselves heard in public life and as part of a functioning democracy. Public libraries could be instrumental in, or take a role in, creating an active democratic communication in matters and issues that are of concern to citizens. However, a task like this cannot be reduced to acquiring and having district plans ready for examination or distributing election campaigning material (flyers, brochures, etc.). It is much more than that. Libraries must dare to act as initiators and take the lead. What must not be forgotten in this respect is the interaction with other media. Consideration should be given to involving several target groups and communities. In the context of the 2009 Municipal Election, video-based profiles and portrayals of the candidates for municipal election were made available. At the same time a group was set up on Facebook in the hope that in relying on this vehicle, there would be better possibilities for appealing to and attracting the interest of younger target audiences.

Digital debate is not better than analogous debate, and you cannot say that it is better to discuss on the web than relying on conventional discussion pieces and letters in newspapers or exchanging questions and views at civic or election meetings. But e-debates facilitated by forums such as the Odder Net in the time before and in the run-up to the municipal election could be instrumental in allowing citizens to make an informed decision when casting their votes. At the same time it is noted that quite a few citizens express themselves only on the Web. Obviously, a certain amount of resources are required for setting up an adequate framework for a debate. Thus, the role of the library is primarily that of a mediator.

6 Concluding Observations

For quite a few years basic public library roles and tasks tended to include such service areas as provision of books and other materials, information services, reference

work, supporting learning activities, organizing cultural activities and promotion of reading. However, during recent years in some countries efforts have been made to redefine public library purpose -the mission of public libraries- with a view to supporting political debates, campaigns, citizens engaging in social and grassroots issues, "activism", etc. But assuming a sharper role in relating to and supporting citizens' political and community-related activities is not a new phenomenon. Actually -as shown by an illustrative case from the Swedish public library history summarized elsewhere in this paper- in some countries there has been a tradition that public libraries committed themselves to making information resources available in connection with community action and citizens' group-based initiatives of various kinds and by hosting discussions and meetings. In this context it is worth referring to the UNESCO Public Library Manifesto, which indicates that the participation of citizens in civic life as an overall aim of public libraries.

The findings and reflections embodied in the Swedish study of the political library and the results of McCabe's analysis of the concept of civic librarianship provide good starting points for further analytic work. In defining an appropriate role for the public library in the Age of Web 2.0, there is a need for reexamining and partly reviving thoughts and ideas on how libraries could support grassroots initiatives and alternative political viewpoints and analyses. Hence, libraries and librarians need to discuss and clarify their stance towards key issues such as participatory democracy, political participation, empowerment and emancipatory roles.

One can imagine that the libraries are keen not to be completely left behind now that e-democracy is taking root in many contexts and environments. Here, the libraries' role can be – as an extension of efforts geared to reduce the digital divide – to provide part of the community dialogue that is undertaken in municipal websites as “real-life” physical sessions (by organizing such activities as politicians' cafés and the like). Still many people do not use or have access to computers and the Internet, and clearly this situation somewhat limits the suitability, performance and impact of Internet-driven social media as a tool of democracy.

The very interesting issue here is: can the public library redefine its mission? Our analytic review of selected readings on Web 2.0 and social media, collective intelligence, digital socialism and the political library has generated some ideas and clues that might be of relevance to the discussion on a changed role for the public library. But can the public library be transformed into an agency that capitalizes on the social media and their innovative applications in supporting democracy, citizen participation in community development and political processes, multiculturalism, etc.? To shed light on this issue more explorative efforts are needed. Thus, in carrying on with the analysis of an updated role for public libraries, it seems obvious to proceed to an empirical study that might be approached as interview-based analysis. For example, a study could be designed that aims to identify selected librarians' views of public library roles in the light of Web 2.0. It would be natural to conduct interview sessions in two or more countries starting with Denmark and Turkey.

In discussing new roles for the public library, there are classic library virtues that should be safeguarded including the library's position as a recognized and trusted repository of information and public knowledge. In a time of booming web technologies and social media and commercialization of information and knowledge there is a need for an agency of neutrality and credibility that helps users unmask the increasing

amount of bias, distortion, fraud, misuse, cheating and manipulation within the fancy new world of web-based media and assists them in navigating in today's information universe, which may be less smooth than imagined. A new user educational perspective would certainly be relevant here.

In analysing the conditions and opportunities for information democracy in the sense of Web 2.0, explorative studies are needed to map politically-related information universes, information transfer and information use. The Digital Age with its new social media invites political engagement, but the era of digitization is also an age of despotic political leader styles, persistent and irremovable power structures, spin doctor-driven politics and infotainment. At least these features seem part of the reality in many countries. Power structures are opaque and various sorts of extra-parliamentary opposition groups, NGOs and grassroots initiatives in specific areas face barriers and difficulties in having their message heard. As is well known, because of failures, backlashes and disappointed expectations situations arise that eventually lead to frustration and apathy. The more than meagre results of COP 15 (United Nations Climate Change Conference) in Copenhagen on the risks and dangers of climate change and global warming come to mind in this respect: about 100,000 committed people walking their planned route in the streets carrying banners and signs and shouting slogans, etc. claiming action on the part of world leaders convening in Copenhagen. They might need a helping hand from libraries.

References

1. Coyle, K.: The Library Catalog in a 2. 0 World. *Journal Academic Librarianship* 33, 289–291 (2007)
2. Stark, E.: *Free Culture and the Internet: A New Semiotic Democracy* (2006), http://www.opendemocracy.net/arts-commons/semiotic_3662.jsp
3. United Nations: *Earth Summit. Agenda 21: The United Nations Programme of Action from Rio*. United Nations, New York (1992)
4. Kenney, G.I.: The Missing Link – Information. *Information Technology for Development* 6, 33–38 (1995)
5. Talero, E., Gaudette, P.: *Harnessing Information for Development. A Proposal for a World Bank Group Strategy*. World Bank Discussion Papers no. 313. The World Bank, Washington, DC (1996), http://www.wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/1999/08/15/000009265_3961219093624/Rendered/PDF/multi0page.pdf
6. Hadden, S.G.: *Democracy on the Electronic Frontier*. In: Chapman, G. (ed.) *Beyond the Endless Frontier*. MIT Press, Cambridge (1996)
7. Doctor, R.D.: Justice and Social Equity in Cyberspace. *Wilson Library Bulletin* 68, 34–39 (1994)
8. Keen, A.: *The Cult of the Amateur: How Blogs, MySpace, YouTube and the Rest of Today's User-Generated Media are Killing our Culture and Economy*. Nicholas Brealey Publishing, London (2008)
9. Eisenberg, M.: The Parallel Information Universe: What's out There and What it Means for Libraries. *Library Journal* 133, 22–25 (2008)

10. Lietzau, Z.: U.S. Public Libraries and Web 2.0: What's Really Happening? *Computers in Libraries* 29, 6–10 (2009)
11. Lancaster, N.: Web 2.0 - Hype or Helpful? [Social Networking: Citizen Engagement]. *Public Library Journal* 23, 6–8 (2008)
12. Luyt, B., Ally, Y., Low, N.H., Ismail, N.B.: Librarian Perception of Wikipedia: Threats or opportunities for librarianship? *Libri*. 60, 57–64 (2010)
13. Brabazon, T.: The Google Effect: Googling, Blogging, Wikis and the Flattening of Expertise. *Libri*. 56, 157–167 (2006)
14. Waller, V.: The Relationship Between Public Libraries and Google: Too Much Information. *First Monday* 14, 9 (2009),
<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2477/2279>
15. McCabe, R.B.: *Civic Librarianship: Renewing the Social Mission of the Public Library*. Scarecrow Press, Lanham (2001)
16. Kranich, N.: Libraries, the Internet, and Democracy. In: Kranich, N. (ed.) *Libraries and Democracy: The Cornerstones of Liberty*, pp. 83–95. American Library Association, Chicago (2001)
17. Astad, C., Curry, A.: Public Space, Public Discourse, and Public Libraries. *LIBRES Library and Information Science Research Electronic Journal* 13 (2003),
http://libres.curtin.edu.au/libres13n1/pub_space.htm
18. Skot-Hansen, D., Andersson, M.: *Det Lokale Bibliotek: Afvikling Eller Udvikling*. Danmarks Biblioteksskole and Udviklingscenteret for folkeoplysning og voksenundervisning, Copenhagen (1994)
19. Matarasso, F.: *Beyond Book Issues: The Social Potential of Public Libraries*. British Library, Research and Innovation Centre, Comedia, London (1998)
20. Durrance, J.C., Pettigrew, K., Jordan, M., Scheuerer, K.: Libraries and Civil Society. In: Kranich, N. (ed.) *Libraries and Democracy: The Cornerstones of Liberty*, pp. 49–59. American Library Association, Chicago (2001)
21. Becker, S., Crandall, M.D., Fisher, K.E., Kinney, B., Landry, C., Rocha, A.: *Opportunity for All: How the American Public Benefits from Internet Access at U.S. Libraries* Institute of Museum and Library Services, Washington (2010),
http://impact.ischool.washington.edu/documents/OPP4ALL_FinalReport.pdf
22. Segaran, T.: *Programming Collective Intelligence: Building Smart Web 2.0 Applications*. O'Reilly, CA (2007)
23. Levy, P.: *Collective Intelligence Mankind's Emerging World in Cyberspace*. Kelly, New York (1997)
24. Kelly, K.: The New Socialism: Global Collectivist Society is Coming Online. *Wired Magazine* (2009),
http://www.wired.com/print/culture/culturereviews/magazine/17-06/nep_newsocialism
25. Jadinge, A.L.: *The Political Library: The Public Library as a space for citizens' participation and public discourse* (2004),
<http://www.abm.uu.se/publikationer/2/2003/189.pdf>
26. Wikinews: *YouTube Banned in Turkey Once Again* (2008),
http://en.wikinews.org/wiki/YouTube_banned_in_Turkey_once_again

Collaboration through Communities of Practice in the Digital Age

Fatih Oguz¹, Corrie V. Marsh², and Cliff Landis¹

¹ Valdosta State University, 1500 N. Patterson St. Valdosta, GA 31698 USA
{foguz, jclandis}@valdosta.edu

² University of Texas - Pan American, 1201 W. University Dr., Edinburg, TX 78539 USA
marshcv@utpa.edu

Abstract. This paper aims to describe and explain the role of Communities of Practice (CoPs) as an informal communication mechanism in initiating, improving, and fostering collaboration in the digital age. CoPs play a critical role in the management of shared knowledge and create value for both their members and organizations. The advent of the Internet and specifically the World Wide Web (WWW) has forever changed the means of accessing and sharing data and information. With the inception of Web 2.0 technologies and social-networking sites in recent years, connections and relationships are now not only nurtured and sustained in an online environment, but also established through creating virtual communities. The authors also assert that the inception of Web 2.0 technologies and social-networking sites is a great advancement in providing a rich learning, communication, and collaborative environment, especially through the transfer of tacit knowledge that we take for granted in our face-to-face interactions. These reflections are based on personal communications with members of virtual CoPs and literature on the impact of CoPs on decision-making and knowledge management.

Keywords: Communities of practice, community informatics, social capital, social networks, social media, virtual communities, Web 2.0 technologies.

1 Introduction

The phenomenon of Communities of Practice (CoPs) has been around for years, and the term itself was first used by Lave and Wenger [1] while studying apprenticeship as a learning model. They argued that the acquisition of knowledge is a social process [2]. CoPs are composed of people who share a concern, common problems, or a passion about a domain, and who want to gain more knowledge and expertise pertaining to that domain through regular interaction [3].

CoPs provide a learning environment through social participation, where participation encompasses much more than engaging in joint activities. Here, it refers to participants being active in the practice and building an identity associated with the CoP to which they belong. Members become aware of their peers' expertise, knowledge, and skills as they engage and interact with each other. They are then able to compare,

verify, and benchmark their professional expertise aligned with their colleagues' knowledge.

In addition, CoPs have the ability to deal with a broad range of knowledge-related issues by connecting isolated professionals with expertise, and linking unconnected activities pertinent to the domain. The individuals who participate in CoPs, as well as the organizations that support and provide resources to them, see value in CoPs for themselves. In the short run, CoPs improve the business outcomes for organizations by providing an arena for problem solving, quick answers to questions, different perspectives on issues, collaboration, and improved quality of decisions. For the members, CoPs may improve their work performance by providing them with access to expertise and knowledge through new ideas and solutions to challenges. However, in the long-run, CoPs develop organizational capabilities by letting organizations envision technological developments and take advantage of emerging market opportunities. For the members, CoPs also foster professional development by helping members to expand their knowledge and expertise in addition to improving their professional reputation [3].

The latest innovations in social media and Web 2.0, are proving to be valuable tools in promoting knowledge creation, dissemination, and preservation. Social-networking sites including LinkedIn, Second Life, YouTube, and Facebook, as well as other interactive Web 2.0 technologies and standards such as Wikis, Blogs, AJAX, and RSS feeds, improve the richness of the information landscape in terms of communicability and interactivity that is traditionally inherent in face-to-face interpersonal interactions. For example, Wikipedia has brought together a community of contributors from around the world through a user-driven Web 2.0 wiki application which was originally developed to facilitate communication among computer programmers in the 1990s. Bejune [4] has identified a number of wikis in the library community to initiate, facilitate, and support collaboration among librarians and their users about various subject areas. Gannon-Leary and Fontainha [5] have noted that virtual CoPs are widely adopted among academics and students as information communication technologies, utilizing e-mail discussion lists and discussion boards since these forms of communication mechanisms have become more user-friendly and interactive.

As commercial organizations expand in size, geographical coverage, and complexity, knowledge has become the key to improving organizational performance. Therefore, the formation of informal social structures like CoPs has become a natural part of organizational life [6], [3]. CoPs make knowledge an integral part of their ongoing activities and interactions. Inter-personal interactions play an important role, especially in sharing tacit knowledge, and the learning tools utilized by CoPs such as storytelling, conversation, and apprenticeship, increase the efficient use of knowledge. CoPs act as a "living repository" for collective knowledge by creating a value for both the members and the organizations supporting and sponsoring these social structures [3].

2 Characteristics of Communities of Practice

According to Wenger [7], a CoP is composed of three crucial characteristics: domain, community, and practice, which together provide a guide to community development and distinguish a CoP from other social structures, such as a project team or neighborhood community.

A domain defines a community through a common framework and identity. It addresses the issues related with a community's purpose such as topics, issues, and benefits pertinent to its members so that a common understanding of the domain can be developed within the community. The domain determines boundaries and guides its members about what is worth sharing and pursuing. It provides them with a direction through which members and other stakeholders are connected to the community. A shared domain encourages members to contribute and participate, and, therefore, provides a sense of accountability to the knowledge that is a distilled product of collective learning.

Wenger et al. [3] define a community as a group of people who engage in joint learning activities, build relationships, and help each other regularly in pursuing their interests in the domain. Continuity in their interactions lets them develop a sense of belonging, identity, and commitment [3]. Nonaka [8] argues that individual commitment is critical to knowledge creation within an organization, since it keeps members engaged in the community affairs. Interpersonal relationships are critical in community building. Knowing "who knows what" [9] makes it easier and efficient for the members to get the 'right answers' they need. In addition, inter-personal relationships enable members to overcome the initial trust issues that may arise when members engage in information sharing activities. Moreover, inter-personal interaction is an effective way of building trust, which is a precondition for genuine knowledge sharing and collaboration [10].

A practice is defined as the set of frameworks, tools, ideas, knowledge, and documents a community develops, shares, and maintains [3]. It refers to the work a CoP's members do and their shared understandings and activities [11]. Moreover, a practice gradually changes as a collective product of a community. The practice is oriented both to past and future. On the one hand, it explores existing knowledge that has been built up and shaped over time by the participants and embodies the history of the community. On the other hand, it looks into the latest advances in the field and thus enables members to handle new situations.

Although a CoP provides its members with a common domain, it does not imply that members have similar backgrounds, skills, and perspectives. A kind of homogeneity may accelerate the community building efforts at the early stages, but it is not a required ingredient for a community. In the long run, continuous interactions among members enable them to build common identity; they also promote diversity. Over time, members develop their own styles and approaches. They define their status within the community by participating in discussions and developing interpersonal relationships. Exemplifying diversity in skills, ideas, and perspectives makes a CoP a richer creative learning environment for its members.

Mutual engagement of members is a personal matter and therefore a source of coherence for the community [12]. From this aspect, participation is voluntary, and it does not really matter how members join, or whether they are self-selected, or

assigned to the community. As they participate in the community activities at various levels, they become the part of the community. Wenger [13], identifies four levels of participation in a CoP as (1) core participants, (2) active participants, (3) peripheral participants, and (4) outsiders.

A small group of people who actively participate in activities of the community often lead and coordinate the community with topics and agendas they determine. The core group constitutes ten to fifteen percent of the whole community. The next level is the active group who attends and participates in the activities but not as regularly as the core group members do. The active group is also small and constitutes fifteen to twenty percent of the whole community. The majority of the CoP members are peripheral and seldom participate in the activities. Some remain peripheral because they think their state of knowledge is not relevant to the rest of the community or carries no authority, or they do not have enough time to contribute. Outsiders are not members of the community but they may have an interest in the community.

Petter, Reich and Helling [14], argue that Web 2.0, specifically social-networking software, can support and facilitate informal learning activities among members of a CoP. Further, they note that widespread availability of the Internet permits like-minded individuals to form virtual CoPs for knowledge sharing activities. Social-networking sites and Web 2.0 technologies drastically reduce the turnaround time necessary for CoP members to gain both explicit and tacit knowledge within a domain. The temporal and spatial limitations of previous forms of communication (e.g., conferences, newsletters/journals, static Web pages) are overcome by a constantly flowing conversation that blends the work and social lives of the CoP members. Members rely on this flowing conversation for professional development and meeting organizational goals.

2.1 Social-Networking Sites

boyd and Ellison [15] define social-networking sites as “web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system.” As the definition suggests, social-networking sites provide a venue for individuals to form networks, not only with their friends, colleagues, and co-workers, but also with strangers. However, such networks generally create weak ties among members, since most members are casual acquaintances and resources and information exchange may be rare [16]. Further, boyd and Ellison [15] argue that a shared offline connection among some of the members, despite these weakly tied relationships, plays an important role in creating virtual social circles.

SixDegrees.com is recognized as the first social-networking site and was founded in 1997 [15]. Since then, the basic functions of such venues have remained the same: to connect individuals based on common interests, friends, or objectives. Today’s social-networking sites rely heavily on Web 2.0 technologies and applications to provide users with more interactive and content-rich environments, and to facilitate the communication processes among their members as seamlessly as possible. Social-networking sites

have policies, procedures, rules, and reporting mechanisms to guide, govern, and control activities of their members and thus to provide a safe online social environment. Nadjm [17] argued that “content moderation” tools such as filtering play a critical role in creating a safe social environment not only for individuals and privacy related issues, but also for organizations through protection of intellectual property.

Ellison, Steinfield and Lampe [18] found that social-networking sites support pre-existing interpersonal offline relations and help individuals keep in touch regardless of geographical and physical boundaries. The networks that are established in social-networking sites are generally formed around people as personal networks [15]. Further, Carter’s [19] findings suggest that personal relationships initiated online can often move to an offline environment and become part of people’s daily lives.

2.2 Social Capital

Social capital can be defined from the CoP perspective as “the common social resource that facilitates information exchange, knowledge sharing, and knowledge construction through continuous interaction, built on trust and maintained through shared understanding” [20]. It is often used as a model to explain various social issues in social groups such as city neighborhoods and is widely discussed in sociology and political science literature [20], [6]. Lesser and Storck [6] argue that social capital is a part of CoPs and gives rise to behavioral changes that, in turn, improve on organizational performance. Their study of communities existing within different organizations (e.g., manufacturing, lending, pharmaceutical) suggests that CoPs reduce the learning curve for new employees by helping them identify knowledge resources within the company, facilitate rapid response to customer inquiries by connecting professionals and expertise, and reduce reinvention by improving reuse of existing knowledge assets. Moreover, they indicated that CoPs serve as generators of social capital by developing, promoting, and nurturing connections and relationships among practitioners regardless of their physical locations and official statuses. In turn, social capital provides a platform where a sense of trust and mutual obligation, shared common language, and context constitute the foundation.

Putnam [21] notes the importance of mutual obligations and reciprocity in social networks and makes a distinction between bridging (inclusive) and bonding (exclusive) social capital. He defines bridging social capital as the loose connections or weak ties among individuals that allow them to access external resources, information and perspectives, and which facilitates the diffusion of information. Bonding social capital is defined as the links between closely connected individuals such as family members or close friends. Ellison et al.’s [18] findings suggest an additional dimension called maintained social capital that explains the ability to stay connected as members of a community move through life. Further, they found that there is a strong association between use of a social-networking site and the three dimensions of social capital: bonding, bridging, and maintained.

3 Collaboration in Practice

The Internet, specifically Web 2.0, redefines, reshapes, and transforms the information and knowledge landscape and how we relate to information and media. Worldwide availability of the Internet and broadband access increase spontaneous encounters and, in turn, virtual CoPs can be easily formed and sustained. Wikipedia (www.wikipedia.org), the biggest multilingual free-content encyclopedia on the Internet, and SourceForge (www.sourceforge.net), the world's largest development and download repository of Open Source code and applications, serve as living products of such encounters and examples of collaborative projects conducted as a community. The popularity of Wikipedia has recently sparked a new Google project called Knol (knol.google.com), a platform for information sharing, that lets users produce knowledge products on any topical area and collaborate with other authors.

Second Life (SL) (www.secondlife.com), a 3D multi-user virtual environment, has served millions of users in a virtual community since its inception in 2003. As in other Web 2.0 applications, residents of SL are able to create their own social spaces and interact with each other. Companies such as IBM have their own SL existence that allow their employees, partners, clients, and other interested parties to meet, learn, engage in business activities, and collaborate with each other. Libraries, museums, and educational institutions from all over the world are also trying to leverage the functionalities offered in this online environment by building collaborative learning environments [22], [23], [24]. Organizations can create areas such as cafés and conference rooms to allow community members to get together, interact, and improve user engagement.

Oguz [25] found that Web 2.0 applications, such as Wikis, instant messaging, and discussion boards, facilitate building and maintaining communities of practice in academic library digital project collaborations. Further, he noted that CoPs played an important role in enabling staff members to access up-to-date and experienced-based knowledge, provided a distributed problem-solving and learning environment, facilitated informal communication and collaborative activities, and informed the decision-making process.

In early 2005, the authors Marsh and Oguz initiated a CoP in conjunction with their research center activities at the University of North Texas. Marsh, recognizing the synergies between regional radio-frequency identification (RFID) business and local university researchers, originated open forums to introduce potential collaborators. Interests ranged from retail and marketing to information systems, transportation services and security. Local RFID experts from manufacturing, retail, systems, engineering, and security joined the forums for presentation and open discussions, resulting in collaborations and projects. An original group of around ten grew to a network of over fifty participants and approximately eight sub-groups. During this process a virtual knowledge base is created for capturing conferences, presentations, and resources to create and promote an online community identity. This mix of personal and web-based contact provided faster communications and knowledge exchange leading to a high level of internet discoverability.

3.1 A Virtual Community of Practice: The Library Society of the World

The Library Society of the World (LSW) (thelibrarysociety.org) is an informal community of librarians that formed to provide community and support outside of the confines of the American Library Association (ALA). As founding member Joshua M. Neff describes.

A bunch of us library types were chatting on Twitter one day, complaining about the ALA. My chief complaint was (and still is) that I can't actively participate in the ALA, because it costs too much (basic membership plus joining divisions or round tables plus attending in-person meetings) compared to what I get paid. Someone, I can't remember who, suggested that internet technologies should allow us to create our own grassroots library association. Someone else issued a dare to start something. Inspired, I came up with a name I liked (being a long-time fan of superhero comics like "Justice Society of America"), quickly created a logo and created a wiki. Then I sent out a link to the wiki on Twitter and let word of mouth spread from there (personal communication, August 24, 2008).

And word of mouth spread quickly: the loosely connected CoP of techie librarians (technology savvy) soon gathered around the principles of an informal and often humorous approach to collaboration. This informality displays a degree of trust found in social capital. Coleman [26] argued that "social capital is created when the relations among person change in ways that facilitate action... a group whose members manifest trustworthiness and place extensive trust in one another will be able to accomplish much more than a comparable group lacking that trustworthiness and trust." Much of the language used to define both CoPs and social capital is echoed in the experience of Iris Jastram, another founder of LSW:

We saw a need for a grassroots community to work together to support each other professionally, provide point-of-need professional development for each other, and develop social ties together. We were able to fill that need because the original community gelled at a time when key social tools online were mature enough to make this kind of blended professional and social interaction effortless. And I'd like to emphasize (again) the blend of professional and social interaction that has always been a fundamental part of what we do. Take the social element out and we wouldn't have the trust level to lay bare our ignorance and ask "silly" questions. Take the professional element out and it would fall by the wayside as we got busy with our day-to-day jobs. The blend is key (personal communication, August 26, 2008).

This informality and the relaxed approach to community highlight much of what members were looking for. Members' comments suggest that as the LSW developed, weak ties between members became stronger as the bridging social capital increased. Members are able to actively engage in discussions and contribute to the LSW because of its foundation in Web 2.0 technologies. Without the initial conversations among techie librarians using Twitter (www.twitter.com), a micro-blogging tool, the idea for the community might never have happened. After the initial conversation, a

wiki was created to capture and codify the conversation that was happening among members. This was followed by a Meebo room (www.meebo.com/rooms) where members could chat freely and later other social-networking sites and Web 2.0 technologies (e.g., interactive discussion forums using Tangler (www.tangler.com), Flickr (www.flickr.com) and LibraryThing (www.librarything.com) groups) to keep the members better connected and capture and disseminate knowledge generated in the community. With communication options being spread across several Web 2.0 technologies, the best applications became those that were used the most. As one member describes, “As an LSW branded thing, I’d say the Meebo room has been the most rewarding for me. Just by having the occasional conversation there with colleagues, I would say I have indirectly helped about 50 different libraries and thousands of library customers. In return, I would have received an equal 50 helps back” (personal communication, August 27, 2008). This serves as an example of the reciprocity found in the social capital as described by Putnam [21] and reflects upon two important characteristics of CoPs, sense of belonging and mutual engagement, as Wenger et al. [3] described.

Members of a CoP often recognize the difference between traditional bureaucratic organizations and informality in interactions in CoPs. As another member states:

[M]y initial understanding of LSW was that it would be an alternative to other professional organizations. But it's so different from every other professional organization I've been involved with that I think it's unfair to call it an alternative. To me, it's become a new way of communicating with and learning from my colleagues. It's less formal, yes, but I feel like I have richer interactions more frequently because of it (personal communication, August 25, 2008).

However, despite its informal atmosphere and short age, the LSW has already served as a forum for professional collaboration. In addition to the creation of a collaboratively managed website and blog, core members of the CoP have given professional presentations on the LSW’s formation and growth, and a free, online professional development conference is being planned.

Web 2.0 technologies are making it possible for these communities to reach a broader population, thereby increasing the social capital of their members. Where traditional professional organizations have been bound by geography, discipline, or the ability to pay dues, CoPs like the LSW are being born online, reaching out to members across the globe, and encouraging a discussion of social and domain-specific topics that benefit their members.

3.2 The Community Pulse: Friendfeed

Not all CoPs are as formal as the LSW. Many of the techie librarians who are members of LSW also socialize with peripheral members and outsiders via Web 2.0 technologies. One social-networking site that has played an important part in this broader CoP is friendfeed (friendfeed.com). friendfeed is a social aggregator—users can connect the accounts that they hold on various social websites (e.g., YouTube, MySpace, Facebook, del.icio.us, Twitter, LinkedIn, and Flickr). Once these accounts are linked, the user and his or her friends will see a stream of that user’s activity from each of

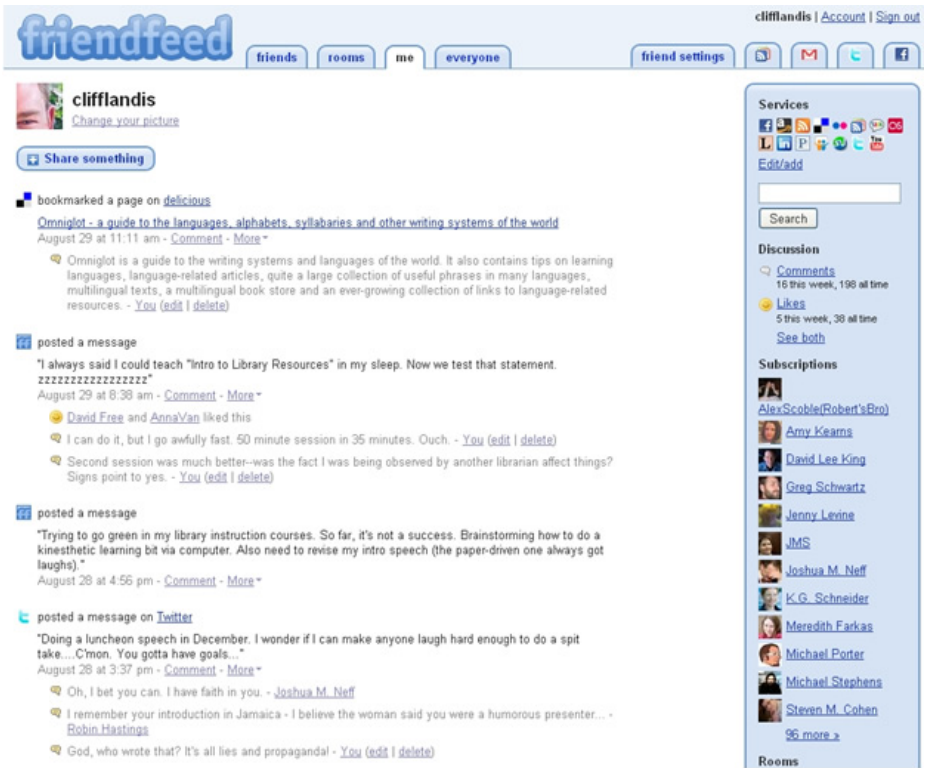


Fig. 1. Professional and Social Interaction on friendfeed

these separate accounts on friendfeed. Additionally, users can like a post (showing interest/approval), and they can also comment on each others' online activity, which allows for both professional and social interaction as shown in Figure 1.

The possibility for CoP members to engage each other in conversation about their online activity can have a profound impact on the professional development and social capital. For example, a recent message reflects on the impact of social media: "Social software has brought about an entirely new aspect of the psyche: the extra-ego, or hyper-ego. I have outsourced part of my psyche to a self-selected peer group, which acts as a validator, gut-check, and willing audience for a selection of my accomplishments and foibles" (personal communication, August 25, 2008). This post not only illustrates the nature of a virtual CoP and the impact of social capital on its members, but it also demonstrates the role of a social-networking site as an information sharing platform.

As Web 2.0 applications such as friendfeed have emerged, collaborative behavior typically found in a CoP has been visibly altered. Traditional collaborative practices such as information seeking, discovery, dissemination and evaluation have been freed from the previous limitations of geography. CoP members can now seek the input of domain experts in real-time, regardless of their organizational affiliation or location

on the globe. In addition, the turn-around time for domain communication and collaboration has been significantly reduced. This ongoing conversation contributes to members' professional growth and development. As one member states, "I love how much I learn on friendfeed, and how much I can bring back to my own library. I don't get out to conferences a lot, but I almost feel as if I did, because there's so much to learn from colleagues on the social web" (personal communication, August 25, 2008). This ability of virtual CoPs to overcome the barriers that slowed communication for face-to-face CoPs will have a strong impact on traditional organizations and their members in the future.

4 Learning and Knowledge Processing in CoPs

In order to examine the benefits and challenges of CoPs it is important to explore the process of knowledge creation, the nature of human knowledge, and the transfer of knowledge in the CoP domain.

Polanyi [27] categorizes human knowledge as (1) explicit and (2) tacit. The distinction between tacit and explicit knowledge is based on the codability of the knowledge, or the ability to be put into words. Explicit knowledge is easy to specify, document, express verbally and in print, and access. It can be coded and transferred easily via formal channels from one location or organization to another [10]. Tacit knowledge, on the other hand, may not be easily accessible; it is intuitive and experience based. Tacit knowledge cannot be coded and easily transferred [28]; it makes data and fact more meaningful to others who lack that particular tacit knowledge [10]. Tacit knowledge is more actionable knowledge, therefore more valuable as opposed to explicit knowledge [29].

Nonaka [8] proposes that there are four modes of knowledge conversion: (1) from tacit knowledge to tacit knowledge (socialization), (2) from explicit knowledge to explicit knowledge (combination), (3) from tacit knowledge to explicit knowledge (externalization), and (4) from explicit knowledge to tacit knowledge (internalization). His model is based on the assumption that "knowledge is created through conversion between tacit and explicit knowledge" (p.18). Although tacit knowledge cannot be articulated or easily put into words, a shared repertoire developed in CoPs facilitates communication of tacit knowledge.

Gannon-Leary and Fontainha [5] list several success factors, including the sense of belonging and trust, within virtual CoPs. They also delineate several challenges that serve as barriers to virtual CoPs such as cognitive and cultural barriers, legal issues (e.g., intellectual property), and lack of face-to-face interaction. In addition, anonymity, or hidden identities in an online environment, is cited as another potential barrier that prevents members from getting to know each other and build trust, both of which are critical to information sharing. Further, participation by a small number of members as opposed to larger number of sideliners, read-only participants, discourages participation. These read-only participants are not uncommon to traditional CoPs in which sixty to seventy percent of the members are peripheral whereas about fifteen to twenty percent of the members regularly participate in community activities and provide direction for the community [3]. It appears that participation-related issues

frustrate those who contribute regularly and have a negative effect on coherence for the virtual community.

As Leadbeater [30] argues, the vast openness of the Web, pervasive use and availability of interactive tools and social-networking on the Internet allow anyone to connect to anyone regardless of organizational or geographic boundaries and facilitate serendipitous encounters, which generates new possibilities for collaboration and helps improve cohesion among individuals.

5 Future Research and Directions

Despite the pervasive impact that the Internet has had since its creation, social media are still a technology in its infancy. Therefore, we are only beginning to see the impact that social media will have on collaborative activities and virtual CoPs over time. Longitudinal studies of CoPs may help us better understand how these communities grow and change over time, especially considering that more and more, thanks to social media and Web 2.0 technologies, these communities are being born online. A further study on the maintained social capital introduced by Ellison et al. [18] may reveal important insights about members' connections in virtual CoPs and its impact on organizational performance. As social-networking sites have gained in popularity, users have been able to reconnect with people from their past. But what will be the impact of relationships that are started and maintained online, throughout the lifespan of CoP members?

6 Conclusion

We now have the ability to create and maintain relationships completely online; relationships that can span the entire distance of the globe with little inconvenience. The advent of Web 2.0 technologies and social media have made it possible for CoPs to establish a new environment for collaboration completely online. This new online environment allows CoPs to reach more new members, regardless of their geographical location, experience level, or ability to financially contribute. It also allows CoP members to communicate and collaborate in less time, and with a broader range. These virtual CoPs allow members to create relationships and develop connections that would have been impossible to establish in the past. These relationships enable the transmission of tacit knowledge between CoP members, bringing context and meaning to the data, facts and numbers that members encounter each day. Each of these abilities has strong implications for both individuals and organizations.

Virtual CoPs' ability to nurture, foster, and transform tacit knowledge provides an unparalleled experience for their members. This ability also makes them an important part of the information environment in today's organizations, as those organizations make the transition to the digital age. CoPs provide their members with a rich and creative learning environment where they are able to gain considerably from diverse skills, ideas, and perspectives available in the community. Engaging in collaborative activities and knowledge sharing are essential to meet organizational goals.

At this point, many members of these virtual CoPs are early technology adopters, and are a minority of the overall population [31]. However, as these technologies become more mainstream, ever more potential members of these CoPs will come online. Even now, we are seeing a strong impact from CoPs that are born digital. It is these CoPs that will serve as both formal and informal communication gateways for collaboration in the years to come.

References

1. Lave, J., Wenger, E.: *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press, Cambridge (1991)
2. Hildreth, P., Kimble, C.: *Knowledge Networks: Innovation Through Communities of Practice*. Idea Group Publishing, London (2004)
3. Wenger, E., McDermott, R., Snyder, W.M.: *A Guide to Managing Knowledge: Cultivating Communities of Practice*. Harvard Business School Press, Boston (2002)
4. Bejune, M.M.: Wikis in Libraries. *Information Technology and Libraries* 26, 26–35 (2007)
5. Gannon-Leary, P.M., Fontainha, E.: *Communities of Practice and Virtual Learning Communities: Benefits, Barriers and Success Factors*. *eLearning Papers* (2007)
6. Lesser, E.L., Storck, J.: *Communities of Practice and Organizational Performance*. *IBM Systems Journal* 40, 831–931 (2001)
7. Wenger, E.: *Knowledge Management as a Doughnut: Shaping Your Knowledge Strategy through Communities of Practice*. *Ivey Business Journal* (2004)
8. Nonaka, I.: *A Dynamic Theory of Organizational Knowledge Creation*. *Organization Science* 23, 14–37 (1994)
9. Cross, R., Parker, A., Prusak, L., Borgatti, S.P.: *Knowing What We Know: Supporting Knowledge Creation and Haring in Social Networks*. *Organizational Dynamics* 30, 100–120 (2001)
10. Persaud, A., Kumar, U., Kumar, V.: *Harnessing Scientific and Technological Knowledge for the Rapid Deployment of Global Innovations*. *Engineering Management Journal* 13, 12–18 (2001)
11. Borgatti, S.: *Communities of Practice* (2004)
12. Wenger, E.: *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press, Cambridge (1998)
13. Wenger, E.: *Supporting Communities of Practice: A Survey of Community-Oriented Technologies* (2001)
14. Petter, C., Reich, K., Helling, K.: *Social Software and the Establishment of Virtual Communities of Practice in the Tourism Sector*. *eLearning Papers* (2007)
15. Boyd, d.m., Ellison, N.B.: *Social Network Sites: Definition, History, and Scholarship*. *Journal of Computer-Mediated Communication* 13 (2007)
16. Haythornthwaite, C.: *Social Networks and Internet Connectivity Effects*. *Information, Communication, and Society* 8, 125–147 (2005)
17. Nadjm, M.: *Security Recommendations for Social Network Communities*. In: *European e-Identity Conference* (2007)
18. Ellison, N.B., Steinfield, C., Lampe, C.: *The Benefits of Facebook “friends”: Social Capital and College Students’ Use of Online Social Network Sites*. *Journal of Computer-Mediated Communication* 12 (2007)
19. Carter, D.: *Living in Virtual Communities: An Ethnography of Human Relationships in Cyberspace*. *Information, Communication & Society* 8, 148–167 (2005)

20. Daniel, B., Schwier, R.A., McCalla, G.: Social Capital in Virtual Learning Communities and Distributed Communities of Practice. *Canadian Journal of Learning and Technology* 29, 113–139 (2003)
21. Putnam, R.D.: *Bowling Alone: The Collapse and Revival of American Community*. Simon & Schuster, New York (2000)
22. Calongne, C., Hiles, J.: Blended Realities: a Virtual Tour of Education in Second Life. In: 12th Annual TCC Worldwide Online Conference (2007)
23. Marty, P., Twidale, M.: *A Second Life for Your Museum: 3D Multi-User Virtual Environments* (2007)
24. McLean, M.: *Library 2.0 and Libraries Building Community Initiatives in Australia* (2007)
25. Oguz, F.: *An Exploration of the Diffusion of a New Technology from Communities of Practice Perspective: Web Services Technologies in Digital Libraries*. Unpublished Dissertation, University of North Texas (2007)
26. Coleman, J.S.: *Foundations of Social Theory*. Belknap Press, Cambridge (1990)
27. Polanyi, M.: *The Tacit Dimension*. Anchor Day Books, New York (1966)
28. Nahapiet, J., Ghoshal, S.: Social Capital, Intellectual Capital, and the Organizational Advantage. *The Academy of Management Review* 23, 242–266 (1998)
29. Marwick, A.D.: Knowledge Management Technology. *IBM Systems Journal* 40, 814–830 (2001)
30. Leadbeater, C.: *We-Think: The Power of Mass Creativity*. Profile Books Ltd., London (2008)
31. Horrigan, J.B.: *A Typology of Information and Communication Technology Users* (2007)

Let Me Tell You Something about (Y)our Culture?

Mícheál Mac an Airchinnigh

School of Computer Science and Statistics, University of Dublin, Trinity College,
Dublin 2, Ireland
mmaa@cs.tcd.ie

Abstract. Each person is born into a culture that is mediated by the mother tongue. Further development of the person is often associated with schooling and education. At an early age some persons will come into contact with other cultures especially if living in a cosmopolitan city or through frequent travel. Such intercultural contact consists of exposure to another tongue, initially aural, and images of the other, perhaps in the form of dress, or architecture, and so on. In the digital world of 2010 those who surf the electronic wave constantly dip in and out of many cultures. Those who normally use Wikipedia in English might over time also refer to a version of an article in another tongue. Those who are frequent users of YouTube might be curious enough to watch a video clip in Turkish or in Greek as well as the usual English, in the context of a history lesson in school. Culture in the digital world needs to be supported and sustained. Are you looking for something? Try Google or Bing or... You have found something you want to share? Post a video clip, or a photograph, or a piece of music. But how shall we keep track of this digital culture? Why would we want to? In this paper we will address the fundamental problem of how to manage cultural information in an integrated fashion in the world of Art. To be specific we will use Bulgarian Art to inform one aspect of Turkish culture.

Keywords: Digital re-discovery of culture, empathy, keyimage, memory, museum of innocence, ontology.

1 Introduction

There seem to be two ways today in which to address a subject: the direct way (preferred in the Sciences), and the indirect way (the choice of Arts and Humanities). However, in most of human endeavour there was always only one way. The (alleged) splitting of that way was introduced to the Anglo-Saxon world by C. P. Snow's *Two Cultures* [1]. In Donald Knuth's [2] Turing Award paper of 1974—"Computer Programming as an Art" [3], he reminds all that the customary word "Ars" for "Art" is the Latin equivalent to the Greek "Techne" from which the word Technology is derived. (Something of this ancient nomenclature survives in the award of the Bachelor of Arts (BA) primary degree to successful undergraduates of Computer Science in Trinity College, by the University of Dublin). This supposed dichotomy between Arts and Science was also reflected in Mathematics, the "modern scientific approach" being the introduction of axioms upon which the subsequent structure was to be built.

Such axiomatic certainty was shaken thoroughly by Thomas S. Kuhn's seminal work: *The Copernican Revolution* [4] and elaborated on further in his essay "Comment on the Relations of Science and Art" [5] where he concurs (with Hafner) that "The more carefully we try to distinguish artist from scientist, the more difficult our task becomes" [5].

In this paper I will try to develop further the idea of the "digital re-discovery of culture" and the game of inquiry. One of the earliest publications in this (new) field is that of Kalina Sotirova [6] where she focused on the idea of "Edutainment", that is to say to be "educated" in an entertaining way. Naturally all education is intrinsically cultural and one of the most enjoyable ways to acquire it is through (game) play. Most languages have just one word for both "game" and "play" [7]. In English, the word "play" is prior. Animals played before humans ever existed. A game is play with explicit rules. In the development of the "game of inquiry" [8], [9], [10] one still seeks to be precise about the rules of play. However, the separation between game and play is not at all clear cut. One speaks of the "Ambiguity of Play" [11]. That is to say, what one thinks is play might be something completely different. This "scientific" paper which follows, is presented in a playful artful manner. And at the very end of the paper, a game of inquiry is presented, as challenge for the reader. So! Without further ado, let the curtains rise and the play begin.

2 Before We Begin in Earnest...

Since the time when more than a century ago... the figure of the so-called Thracian Horseman stood out as the most important diagnostic characteristic of that Thracian heritage... reliefs found (numbering more than 1500 from various parts of Bulgaria...) [12].

When I recently came across this quoted text on the "Thracian Horseman" in the journal article cited, the image of the Madara Horseman [13] immediately sprang to mind. I can not tell if the Madara Horseman¹ is one of these "Thracian Horsemen" implied in the text above. I can tell that I have been to that place near to the village of Madara and have seen for myself, this extraordinary image. Anyone with Internet access can see the image online. And Google Earth will fly you there. It is Google Earth that provides some sort of geographical context — the factness of the plateau rising up out of a flat surrounding plain. To be there in person, to experience the...² And so, from this experience one is led directly to the "not so modern" phenomenon of cultural tourism [14]. One of the consequences of such tourism is often to report back home on what has been seen and experienced. It is not usually the case that one reports back to those whose culture one has experienced. A major stumbling block is

¹ It is noteworthy that the UNESCO World Heritage Centre uses the phrase Madara Rider instead of Madara Horseman. Is this an example of "political correctness"?

² The rest of the text is deliberately omitted. It is assumed that the reader can fill in their own text with respect to a similar sort of cultural experience. For example, one might imagine a reader from Turkey, visiting Ankara for the very first time and being brought to see Anıtkabir. What sort of crossover cultural connections might (s)he then make?

often the natural language³ of the visited culture. And before the “coming into being” of the widespread Web, to what other kind of person might one want to report?

I had seen the Madara Horseman long before I read the journal article. The reading of the article brought back instantly that memory of seeing. It has been and still is customary for us to be acquainted with cultural artefacts, such as sculptures, paintings, manuscripts and so on through the medium of print, often in newspapers and magazines. These latter physical objects have been the stage, the platform, from which each reader breaks out into an imagined world mediated through the print. This leaping out of the imagination from the printed world is at the very heart of print advertising. The book often plays a special role in opening up a culture to another. Historical novels and Memoirs are classic examples. Of particular importance for our “telling about the culture of the other” is the recent work of Orhan Pamuk, *The Museum of Innocence* [15], a work, a novel, that reveals more of the upper class society of Istanbul from which he himself springs. It is a text, a story, a truth revealed, in much the same spirit as his own childhood and teenage years were opened up in the earlier *Istanbul, Memories of a City* [16]. And Pamuk is directly involved in the design and manipulation of the photograph that is used on the cover jacket of the book. There is a car in which three men and two women look at the photographer. But the original background of woodland from somewhere in Turkey’s interior has been replaced with the well-recognized Bosphorus [17]. Pamuk speaks through his translator. Pamuk speaks as one, today, displaced from his cultural roots. The narrator of the *Museum of Innocence* pinpoints the 23-year old chain-smoking Orhan Pamuk as one for whom there is “nothing special about him beyond his propensity to act nervous and impatient, affecting a mocking smile.” [16].

I have been to the street of the Pamuk Apartments, Istanbul, in 2007. I have read all of his books that have been translated into English. I think I know a little of the culture about which he speaks. His works brought me to Istanbul, to Ankara, to Turkey, to understand and empathize with a culture, very different from my own. Through his books Pamuk has implicitly put to me, personally, the rhetorical question “*Let me tell you something about your culture?*” And in this most recent of his novels he puts the question to the “City of Istanbul, European City of Culture”, in July 2010 [18], [19]. The electronic version of the *Museum of Innocence* is already live in Turkish [20] and thoughtfully Pamuk gives the physical location on a map in the book: on the corner of Çukurcuma Cd. and Dalgıç Çk. [21]⁴.

Anyone remotely interested in the politics of civilization will be aware that museums are the repositories of those things from which Western Civilization derives its wealth of knowledge... [15]

Before I read the works of Orhan Pamuk, (we are of the same generation), I already knew Bulgaria, its people, its culture. Today, I realize that I also had begun the

³ Even when the “same language” is involved, such as English, there may be significant cultural differences. A good example might be the difficulty that some American visitors might have after visiting English speaking countries such as Ireland, Scotland, Wales, and even (parts of) England.

⁴ Locating the *Museum of Innocence* via Google Earth took about 30 minutes. I am almost certain that I have the correct coordinates.

construction of my own *Museum of Innocence* in 1979.⁵ Today, in the light of Pamuk's initiative, I think it would be good to mention some of the objects that I will put in my Museum in 2010.

The image of the Madara Horseman is on one side of the 50 stotinki Bulgarian coin [22], from which one may infer that the image is an intrinsic part of the Cultural Heritage of Bulgaria. I have added it⁶ to the Museum collection. I often wonder to what extent the "average Bulgarian" is affected culturally by this image on the coins. Does s(he) learn of its importance through the elementary schooling system? I compare myself with that other and ask to what extent does the "average Irish person" appreciate the meaning of the image of the harp on the reverse of the (formerly) Irish and now European (Irish) coins.

In February 2009 there appeared an article describing a Madara-like Horseman in Shamark, Afghanistan [23]. When I learned of it and looked at the image (on the Internet) I was astonished. I will probably never visit Afghanistan. But the experience of the Madara Horseman already informs me of a little something of the culture of those who made that image. And the 50 stotinki coin currently in my Museum provides a keyimage for the Horseman in Shamark. The images are on the Web. You can compare them to see how alike they are. You too might like to open up your own "Museum of Innocence"?

3 Serendipity

A book, like Pamuk's *Museum of Innocence*, offers an opportunity to the reader to break out into another culture, another era. In earlier times such book breakout was often confined to the physical libraries and museums in the vicinity. All that is now changed by the Web. Let us now turn our attention to works of art, such as sculpture and painting. A sculpture, such as the monumental Lozan Anıtı, Edirne [24] can only be visited or photographed. The photograph of Lozan Anıtı is freely available in many places on the Web [25]. The photograph can never replace the (experience of the) original artwork. But that is true of all such things. A painting stands in a similar fashion. It might be in private hands and, therefore all but the owner and family and friends, are excluded from seeing it. If a photograph or a copy/facsimile/print exists then others may get to experience something of the original. A painting such as Picasso's "Guernica" [26] is on the same scale of inaccessibility as the Lozan Anıtı monument. A photograph or print does not (usually) capture the monumental size of the original work. One might imagine a LED screen technology of the future which would "recreate" "Guernica" in monumental copy form.

In the modern world of social networking [27], whether for personal or professional purposes, it is normal for people to share their "cultural interests". For the "young" such interests are often related to music or film. Photographs will also have

⁵ The year 1979 is chosen to correlate exactly with what the author regards as the opening date for the memory collection.

⁶ By "it" I mean the image of the Madara Horseman photographed. The coin must be returned to the country to which it belongs. And now I realize that I have become an accidental coin collector over a lifetime.

their place. Such “practical sense” suggests a working hypothesis. Philosophically, the medium has changed, not the nature of the interest. So it will be for all. If people do not normally go to Art Galleries to look at certain exhibitions or paintings, then there is no reason for them to do so online. People must be enticed... somehow. To explore this hypothesis let us consider a thought experiment? Let us imagine that you see, for the very first time, a particular picture. Maybe you come across it in a National Art Gallery, such as Dublin or Sofia, or perhaps you see it, by chance, online [28]. Imagine that it is on exhibition in Istanbul in July 2010 to celebrate the year of culture? We summarize these “scenarios of seeing” by placing them in the context of a serendipity hypothesis — the fortuitous accident that opens the door into another cultural world.⁷ A typical invitational text might take the following form:

Let us consider the picture of “The Man in the Red Fez” [28]. What do you see? Can you describe what you see in a few short sentences?

The given title “The Man in the Red Fez” is not the real title of the painting. It is a folksonomical title that anyone might use if (s)he were familiar with the type of hat shown. This particular experiment could only be carried out since mid-2009 when the image became available online. Here is the background scenario designed to focus on a particular aspect of culture.

In our modern times (2010) the deliberate choice of the use of the word Fez in the title of the painting, rather than Hat, will in all probability suggest to the curious one, the desire or need to a) search with Google, b) search Wikipedia, or even c) search “an online dictionary” such as Wordnik [29]. The Wikipedia article gives a focused though ambiguous result: 11 possibilities, of which the 3rd brings one to the hat interpretation. A subsequent quick perusal of the Wikipedia text will lead one directly to the text:

“In Turkey, wearing the fez was legally banned in 1925 as part of the modernizing reforms of Mustafa Kemal Atatürk.”

That picture of the man in the red fez⁸ is now hanging in a prominent place in my museum.⁹ The curious will wonder why the fez was banned. Tourists can still buy the fez in Istanbul, for example. But it is not the real thing¹⁰ [30]. The painting “The Man in the Red Fez” is a keyimage. It unlocks the door to a very special place in Balkan and Turkish history. What kind of place might that be?

In “my picture of the world”, the banning of the fez brings to mind immediately some very beautiful portraits of Mustafa Kemal Atatürk [31] which are on display in

⁷ Such a cultural world might very well be one’s own youthful cultural environment. I remember when I was 13, and I remember walking out on the ice-covered lake. One year later I would leave home forever. This year, writing from another place, in 2010, has the feeling of 1963, all over again.

⁸ The emphasis on “red fez” is deliberate. If used on its own, it is assumed that the fez is red. However, other colours of fez are in current use. There are pictures of Mustafa Kemal where he appears to be wearing a black fez.

⁹ <http://www.flickr.com/photos/mihalorel/4268360773/>

¹⁰ Where did I read this recently (January 2010)? Was it in a newspaper? Did I read it online? Sometimes one does forget to mark the source. When I was in Istanbul in 2007, I do not recall seeing a fez for sale.

the War of Independence Museum [32]: captioned¹¹ “Gazi Mustafa Kemal riding the horse¹² called “Sakarya” (undated), “Commander-in-chief Marshal Gazi Mustafa Kemal” (1922) and “President Gazi Mustafa Kemal” (1923). In each of these pictures a black hat of a certain form is worn. It is of a different style to the fez. I wonder about the name and origins of this black headpiece. And for the “modern generation” let us introduce a YouTube clip showing some very old film which gives one some further idea of the headdress styles around 100 years ago [33].

Let us return to the painting “The Man in the Red Fez” and its location. The visitor to the art gallery in question (whether in Sofia or online) will see that the man in the red fez is a painter by the name of Georgi Danchov [28]. A search in Wikipedia will reveal that a) there is no page in English for him, and (b) he was an associate of Vasil Levski.¹³ There is, of course, a page for Георги Данчов in the Bulgarian Wikipedia. And every Bulgarian knows the significance of Vasil Levski, a person on a par with Mustafa Kemal Atatürk for those of Turkish culture.

Although the wearing of a Fez is not and never has been a part of “my” culture, I have a “feeling for” those cultures in which it was (and still is). Part of that feeling is not only visual but one of feeling, experienced through music [34]. That is to say, I listen to the latter as I am writing and remembering. And remembering the year of the Fez ban (1925) brings one back to that time when the centre of power shifted from Istanbul to Ankara. And from a Culture point of view one might say that today in 2010 there is a refocusing of “the” culture in Istanbul [18]. In today’s electronically connected world one can get a sense of the fire of Anatolia on YouTube [35], and the experience will trigger remembrance of Riverdance among people of “modern Celtic culture) [36]. Ten years after Eurovision in Dublin, Istanbul opened with Sertab Erener, and at the end of her performance (part 2) a well-known hat appeared, not a fez but the headdress of the Sufi [37].

Let us conclude this section by explicit reference to the kinds of headdress worn by males and the cultural significance of same. Males in the army and in the police also use distinctive headdress. Under current equality legislation in many cultures, females wear the same sort of headdress as the males, an interesting development in our times. In some cultures, the male headdress can be very provocative. One classic example is that of the Sikh’s turban [38] which came into conflict with the headdress to be worn by the police. Another provocative situation arises in the wearing of the “bowler hat” in Northern Ireland [39].

4 Keyimage and Ontology

From an information management perspective we shall find a (personal) need to classify what we have unearthed from this self-portrait of Georgi Danchov and his

¹¹ I have no information on the name of the painter who executed these works. How might one find out on the Web?

¹² Deliberately chosen to reflect back to the Madara Horseman. The significance of the Horseman (Centaur) is deeply rooted in certain cultures.

¹³ The omission of reference material to Vasil Levski is deliberate in view of the game yet to be played.

fez. It is our need, precisely because we have discovered this chain of thoughts and events from a particular source: a picture, something which we will call a keyimage. And it is the Red Fez in the picture which makes the latter a keyimage. From the point of view of the CIDOC-CRM [40], the object “Red Fez” which is portrayed may be formally classified as an E22 Man-Made Object. But surely this real “Red Fez” object of 1867 no longer exists? If it does then we may formally record certain properties that it possesses, such as P54 has current permanent location, and so on.

The building up of large-scale relevant ontologies is a massive and competitive task. There are professional bodies for which such ontology-building is their “bread-and-butter”. Consider, for example, the Getty Research unit [41]. Let us look to see if they have any record of a famous Bulgarian artist¹⁴ such as Ivan Milev [42]? Using the GRI Catalog Plus (BETA) [2010-01-16] returns 0 results in all categories. On the other hand, a search in their Union List of Artist Names [2010-01-16] gave 1 result, the one desired. Curiously, a search of the same list with Иван Милев, led to <http://www.getty.edu/redirect/500.html?error=java.lang.ArrayIndexOutOfBoundsException>, and an on-screen message: “*We're sorry — we are currently doing maintenance on the site. Please try again in a few minutes.*”

Whatever process is used by Getty Research, it seems to me to be clear, at this time, that they are not “scraping” Wikipedia pages (in any language) for possible candidates to add into their databases.

Page scraping by Web spiders is currently one of the practical ways to harvest data. What Google does, works extremely well, and to a certain extent depends entirely on what we humans contribute as data. Tim Berners-Lee already expressed this idea [43], suggesting that “documents [...] created within the same logical framework, such as RDF, [will lead to] partial understanding [and that] This is how computers will work across boundaries, without people having to meet to agree on every specific term globally.”

We can explore the state “of the art” (pun intended) by choosing some “official pages” of Art Galleries and viewing the source. For “our man” with the Red Fez, the relevant code is

Table 1. Information gleaned from Web Page

<code><table width="1" height="1" border="0" cellpadding="4"></code>
<code><tr><td style="border: solid #DCDFE6 3px;" ></code>
<code></code>
<code></td></tr></code>
<code></table></code>
<code>

</code>
Автопортрет, 1867
<code>
</code>
Георги Данчов (1846-1908)

¹⁴ We use Ivan Milev rather than Georgi Danchov to illustrate the use of the Getty research tools. Georgi Danchov does not even turn up in the Union List of Artist Names.

The corresponding English version differs only in the last three rows:

Table 2. Corresponding English text

Self-portrait, 1867
Georgi Danchov (1846-1908)

There is no mention of a “Red Fez”. That is not surprising. But the very fact of the absence of “Red Fez” or “Червен фес” means that this picture will not be found directly on its home web site if we use these search terms. This representation of a work of art is not unusual. In his paper “How formal is the structure of historical knowledge?” Martin Doerr highlighted the fundamental problem in a keynote presentation on the CIDOC-CRM in the First International Workshop [on] Ontology based modeling in Humanities:

All information fields [of things in the museum Benaki [44] in Athens] appeared as if they were parts of the object itself and not entities in their own right, such as “creator” and “creation date”, and uncorrelated between each other in a natural way. [45]

We note that this problem also pertains to the painting which we have called “The Man in the Red Fez”. The painter is listed as “Georgi Danchov (1846-1908).” As humans we know that Georgi is the first name, that Danchov is the family name. We infer that this is the name of the painter. This supposition is confirmed by the title of the painting: “Self-portrait.” The date range is understood to give his birth: 1846 and his death: 1908. Finally, the painting is dated 1867. If we add in the (original) Bulgarian text for the same painting we, as humans, may deduce that “Георги Данчов” corresponds exactly to “Georgi Danchov.”

5 Ontological Detail

Before the existence of the CIDOC-CRM I used a simple ad-hoc ontological structure for research into Cultural Heritage, with special focus on Art. In particular under the Class Person, I used subclasses Artist, Owner, and Photographer. The Artist class was further subdivided into Painter and Sculptor. For the Class Genre, I restricted myself to History Painting, Landscape, Nude, Portrait (with subclass Self Portrait) and Still Life. There are some interesting problems associated with the unique identification of a Painting, for example. Consider the portrait of Atatürk in 1923.

Introduction of the CIDOC-CRM required that my Person class be made equivalent to the E21 Person. This sort of action goes by the name of ontological alignment [46], [47].

For the class Person, the properties of birth date [P981 was born], death date, and death place were regarded to be fundamental. Another property, unique to the author, was that of a keyimage. For the painting in question, “The Man in the Red Fez,” the most important thing to do is to give it some sort of unique identity. Doubtless the National Art Gallery in Sofia has such an identifier, required by Bulgarian law, and inscribed both on the Passport as well as the Dossier of the painting. Unfortunately, I

am not currently privy to that information. For the purposes of research, a unique name can be assigned within the framework of Flickr. Currently, a version of "The Man in the Red Fez, is located at the unique address:

<http://www.flickr.com/photos/mihalorel/4268360773/>

This address is re-encoded for the CIDOC-CRM OWL ontology within Protégé in the form `Place_Flickr_mihalorel_4268360773` and annotated using Dublin Core [48] metadata:

Source: <http://www.flickr.com/photos/mihalorel/4268360773/>

Table 3. Dublin Core metadata for "The Man with the Red Fez" painting, 1867

contributor	(An entity responsible for making contributions to the resource.) Михал Орела
coverage	(The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.)
creator	(An entity primarily responsible for making the resource.) Георги Данчов (1846-1908)
date	(A point or period of time associated with an event in the lifecycle of the resource.) 2010-01-16
description	(An account of the resource.) The resource is considered to be that photograph of the original painting (self portrait) of Georgi Danchov, 1867. The photograph in question is located on Flickr with URI http://www.flickr.com/photos/mihalorel/4268360773/
format	(The file format, physical medium, or dimensions of the resource.) The format is jpeg: http://farm5.static.flickr.com/4015/4268360773_caac7f3e33_o.jpg and has dimensions width="500" height="706"
identifier	(An unambiguous reference to the resource within a given context.) http://www.flickr.com/photos/mihalorel/4268360773/
language	(Recommended best practice is to use a controlled vocabulary such as RFC 4646 [RFC4646].) en
publisher	(An entity responsible for making the resource available.) Михал Орела, Flickr.
relation	(A related resource.) http://www.nationalartgallerybg.org/index.php?l=55&id=28
rights	(Information about rights held in and over the resource.) Since the resource referred to is on Flickr, then one is referred to http://creativecommons.org/licenses/by/2.0/
source	(A related resource from which the described resource is derived.) http://www.nationalartgallerybg.org/index.php?l=55&id=28
subject	(The topic of the resource.) painting, oil on canvas, self-portrait, Georgi Danchov
title	(A name given to the resource.) Автопортрет 1867
type	(The nature or genre of the resource.) digital photograph

6 Comment and Assessment

It is quite tedious to produce Dublin Core (DC) metadata by hand. Moreover, notice that the 15 elements are listed in alphabetical order. This is also the order used in the CIDOC-CRM OWL ontology running in Protégé 4. Machines do not really care what order metadata elements are listed. From a human perspective, the DCdot tool [49] may be used. Let us look at another portrait by the artist Georgi Danchov on the Web [50]. Scraping the page for DC metadata using DCdot gives the following result:

Table 4. DCdot metadata for Newspaper Вестник "Марица"

Title	Гледаме уникати на Георги Данчов-Зографина - Вестник "Марица"
Creator (author)	
Subject or keywords	SMS; 10:52; Martin; 19:37; 11:20; 10:39; Sport; 19:07; 14:31; 18:30; 19:40; 20:39; Diesel; maini4ka_1; btgeorgiev; 16:09; 20:08; 22:38; Google; Viscomp; niki8283; 21:44; 14:24; Maps; 14:09; 21:48; 12:26; Arruor; Facebook; 0897/577-346; cunitooo; Delicious; 17:15; beninda; Digg; gorgona; Mercedes; 08:27; ogi_pld; Svejo.net; kasteli; 02:19; 21:26; 12:04
Description	
Publisher	
Contributor	
Date	
Type	Text DCMI Type:
Format	text/html; charset=UTF-8l 57465 bytes
Identifier	http://www.marica.bg/show.php?id=9047
Source	
Language	
Relation	
Coverage	
Rights management	

It is the author's experience that DC metadata is often quite sparse for Newspapers online. It is also quite common to find that the Subject or keywords box is very large.

It is the author's opinion that the quality of the DCdot data is often a good indicator of the online structure of the Newspaper in question. Hence one may say that the DCdot tool measures to a certain extent just how well the much mooted ontological infrastructure has actually penetrated in practice.

7 Digital Re-discovery of Culture and the Game of Inquiry

Very few people will ever visit the National Art Gallery (NAG), Sofia, and study the painting of "The Man with the Red Fez". Why would anyone want to? But now that there is a digital version of the painting on the NAG web site and, more importantly another digital version on Flickr, then we have the possibility of telling a story, many different stories, in fact.

One such story might begin with a simple observation:

Once upon a time, there was a painter who liked to dress up. He was a very good painter. One day he said to himself, I think it is time to paint myself. I am 31 years of age. I am in the prime of life. I will put on my very best clothes and I will wear my lovely red hat.

Notice that there is no mention of the date 1867 and "red hat" is used instead of "red fez". Nor do we mention the fact that the painter is Bulgarian. To give very specific details of geographical and historical fact in a story will already bias the story. The full version of this story is to be used as a background, a stage setting, for a kind of play, a piece of theatre, a game.¹⁵

The next step in the preparation of our play will be the introduction of some scenery, some stage props. Why not introduce the Flickr address which gives access to the self-portrait of Georgi Danchov? Alternatively we can give the Wikipedia address of the page on which a "similar" self-portrait appears. This particular portrait is a key-image in the play. It shows a well dressed man with a red hat. We might choose another image of a man with a similar looking hat from around the same era. What about the photograph of a pair of Bulgarian Turks from "село Бояна (Кокарджа) варненско" [51] This old photograph is black and white. We can only assume that the man on the left is wearing a red fez. Put the two images side by side and it is clear that the fez was worn both by elegant painters and more humble peasants. Such juxtaposition of two "opposing" images raises a question in the mind of the player of this "game".

In the English language, the notion of a game requires a set of rules and, in addition a space in which the game is played [52]. The space for the game of inquiry is the Web, augmented if necessary by one's own physical resources. For example, the game being designed here requires the use of books and other physical things, such as prints of paintings. A reasonable account of the structure of the Game of Inquiry is available [8]. Here we conclude with yet another summary attempt at describing the structure, the paraphernalia,... of the game of inquiry. In brief, there are 5 key elements: 1) the backstory (max 1 A4 page) which sets the scene for the

¹⁵ In many languages there is no distinction between the word for play and the word for game. In Bulgarian there is just the one word ИГРА. The author is playing a game with words deliberately.

drama to be enacted; 2) the set of keyimages (augmented with video clips, music clips, sound clips) which provides the scenery, the backdrop, the feeling of space, the mood of the piece; 3) the set of keywords or keyphrases that help to focus the direction in which the player makes a move; 4) the set of Web pages, each of which is a window onto a bigger relevant landscape; 5) the explicit stated goal of the game. Items 2 and 4 need to be kept low in numerical number. In particular, the use of Web pages as a breakout device is fraught with the possibility of delivering a “never ending story” type of game. Let us conclude the discussion with a game based on materials already presented? And what shall be the subject? Given the factness of Istanbul as EUropean City of Culture 2010 and the opening of the Museum of Innocence there in July, then it does seem to be appropriate to devise a game “out of” Orhan Pamuk’s most recent novel. Space permits but a brief outline of the game:

Backstory: “Mr. Orhan Pamuk recalled that Füsün had danced with two people early on. He didn’t know or couldn’t remember her first dance partner... The second, however, was the young man...—Orhan Pamuk himself... Those interested in Orhan Bey’s own description of how he felt while dancing with Füsün should look at the last chapter, entitled “Happiness.” [15]

Keyimages: 1) [53]; 2) [54]; 3) [55]

Keywords: hotel, engagement party, flying carpet canopy

Web pages: 1) [56]; 2) [19]; [57]

Goal: Where did Orhan Pamuk dance with Füsün?

Acknowledgments. I am thankful to Kalina Sotirova for the reference to the Madaralike Horseman in Afghanistan. Thanks are due also to Kalina and my daughter Thérèse for proof reading an earlier draft of the paper. A very special note of gratitude is due to the two anonymous reviewers whose insightful comments removed the scales from my eyes, showing me that perhaps I presumed too much.

References

1. Snow, C.P.: *The Two Cultures*. Cambridge University Press, Cambridge (1998)
2. Wikipedia, http://en.wikipedia.org/wiki/Donald_Knuth
3. Knuth, D.: *Computer Programming as an Art*. *Communications of the ACM* 17, 667–673 (1974)
4. Kuhn, T.S.: *The Copernican Revolution; Planetary Astronomy in the Development of Western Thought*. Harvard University Press, Cambridge (1957)
5. Kuhn, T.S.: *The Essential Tension: Selected Studies in Scientific Tradition and Change*. University of Chicago Press, Chicago (1977)
6. Sotirova, K.: *Edutainment (Game) – Digital (Re)Discovery of Culture*. *Review of the National Center for Digitization* 6, 58–68 (2005)
7. Huizinga, J.: *Homo Ludens; A Study of the Play-Element In Culture*. Beacon Press, Boston (1955)
8. Mac an Airchinnigh, M.: *The Digital Culture*. Paper presented at the Balkan Ülkeleri Kütüphanelerarası Bilgi-Belge Yönetimi ve İşbirliği Sempozyumu (2008)
9. Mac an Airchinnigh, M., Sotirova, K.: *Digital MultiCulture in Practice*. In: Hahn, W., von Vertan, C. (eds.) *First International Workshop Ontology Based Modelling in Humanities Bericht*, vol. 264, pp. 27–36. University of Hamburg (2006)

10. Mac an Airchinnigh, M., Sotirova, K., Tonta, Y.: Digital Re-discovery of Culture Game of Inquiry & the Physicality of Soul. Review of the National Center for Digitization 9, 19–37 (2006)
11. Sutton-Smith, B.: The Ambiguity of Play. Harvard University Press, Cambridge (1997)
12. Toporov, V.N.: The Thracian Horseman in an Indo-European Perspective. Orpheus, Journal of Indo-European and Thracian Studies 0, 46–63 (1990)
13. UNESCO World Heritage Centre, Madara Rider,
<http://whc.unesco.org/en/list/43>
14. Wikipedia, http://en.wikipedia.org/wiki/Cultural_tourism
15. Pamuk, O., Freely, M.: The Museum of Innocence 1st North American edn. Alfred A. Knopf., New York (2009)
16. Pamuk, O.: Istanbul: Memories of a City. Faber and Faber, London (2005)
17. Quinn, B.: Pamuk Photoshops. Review Saturday Guardian. 230110, 5 (2010)
18. Hurriyet Daily News and Economic Review,
<http://www.hurriyetdailynews.com/n.php?n=istanbul-launches-year-as-european-cultural-capital-2010-01-17>
19. Istanbul 2010 - European Capital of Culture (2010),
<http://www.en.istanbul2010.org/index.htm>
20. Masumiyet Müzesi, <http://www.masumiyetmuzesi.com/public/>
21. Google Earth. Museum of Innocence, 41° 01' 50.72" N 28° 58' 47.06" E
22. Мадарският конник, Мадара,
<http://www.flickr.com/photos/mihalorel/4304720466/>
23. Мадарски конник в Афганистан. Списание *8* - за твоя космос,
<http://www.ossem.eu/article.aspx?pg=ar154>
24. Lozan Anıtı, Edirne,
<http://www.flickr.com/photos/mihalorel/2563847728/>
25. Lozan Anıtı, http://www.trakya.edu.tr/Koruma/Lozan_Aniti.htm
26. Guernica,
<http://en.wikipedia.org/wiki/Image:PicassoGuernica.jpg>
27. Krotoski, A.: The Virtual Revolution, How 20 Years of the Web Has Reshaped Our Lives (2010)
28. Georgi Danchov (1846-1908), Г. Д.: Self-portrait (Автопортрет) (1867)
29. Wordnik, <http://www.wordnik.com/>
30. van der Lit, M.: Red hat for your head. (Grand Bazaar, Istanbul, Turkey) (2008),
<http://www.flickr.com/photos/bredero5/2633865053/>
(retrieved January 31, 2010)
31. Mustafa Kemal Atatürk,
http://en.wikipedia.org/wiki/Mustafa_Kemal_Ataturk
32. Genelkurmay Başkanlığı, T.C.: Atatürk ve Kurtuluş Savaşı Müzesi (Ataturk and the War of Independence Museum). Ankara
33. Mustafa Kemal Ataturk Paşa,
<http://www.youtube.com/watch?v=avmRFsGoeKQ&feature=related>
34. Anadolu Ateşi (The fire of Anatolia): Yeni Dünya Müzik, İstanbul (2004)
35. Fire of Anatolia@Eurovision Turkey—Full Version (from German TV),
<http://www.youtube.com/watch?v=1709-M9jZ10&feature=related>
36. Riverdance@Eurovision (1994),
<http://www.youtube.com/watch?v=jWyhH5z92ps>
37. ESC Eurovision 2004 TURKEY—İstanbul Opening,
<http://www.youtube.com/watch?v=ICTx9ILlqBM&NR=1>

38. Sikhi Wiki. Turban, <http://www.sikhiwiki.org/index.php/Turban>
39. Demotix Bowler Hats, Sashes and Banners: the Orange Order in Northern Ireland.(2010), <http://www.demotix.com/news/30876/bowler-hats-sashes-and-banners-orange-order-northern-ireland> (retrieved January 24, 2010)
40. Schiemann, B., Oischinger, M., Görz, G.: Erlangen CRM Version 2009-03-30 (2009), http://www8.informatik.uni-erlangen.de/IMMD8/Services/cidoc-crm/erlangen-crm_090330_5_0_1.owl (retrieved April 15 2009)
41. The Getty Vocabularies, http://www.getty.edu/research/conducting_research/vocabularies/
42. Ivan Milev, http://en.wikipedia.org/wiki/Ivan_Milev
43. Berners-Lee, T., Fischetti, M.: Weaving The Web: The Original Design and Ultimate Destiny of The World Wide Web By Its Inventor. HarperCollins Publishers, New York (2000)
44. Museum, B.: <http://www.benaki.gr/index.asp?lang=en>
45. Hahn, W.v., Vertan, C.: International Workshop Ontology Based Modelling in Humanities (1 2006 Hamburg). In: First International Workshop Ontology Based Modelling in Humanities, April 7 - 8. Univ. Hamburg, Bibliothek des Departments Informatik Hamburg (2006)
46. Lockheed Martin Corporation, <http://www.atl.external.lmco.com/projects/ontology>
47. Ontological alignment, http://en.wikipedia.org/wiki/Ontology_alignment
48. DCMI. Dublin Core Metadata Initiative, <http://dublincore.org/>
49. DC-dot, <http://www.ukoln.ac.uk/metadata/dcdot>
50. Вестник "Марица". Вестник "Марица", <http://www.marica.bg/show.php?id=9047>
51. Български турци от село Бояна (Кокарджа) варненско, облечени в типични носии на заможни, млади мъже, с бели пояси и богато подшити с гайтани горни дрехи - около 1880-1890г, <http://www.lostbulgaria.com/?p=1900>
52. Salen, K., Zimmerman, E.: Rules of Play: Game Design Fundamentals. The MIT Press, Massachusetts (2004)
53. El Museo de la Inocencia, <http://www.flickr.com/photos/cargadadelibros/4288340993/>
54. Pamuk Apartment, Nisantasi, Istanbul, <http://www.flickr.com/photos/69331295<Literal>@N00/2438784870/</Literal>>
55. Hilton, I.: http://www.diploma-unit-9.net/Omer/2008/11/hip_hotel.html#more
56. European Capital of Culture, http://en.wikipedia.org/wiki/European_Capital_of_Culture
57. Oteli, İ.H.: <http://arkiv.arkitera.com/p2247-istanbul-hilton-oteli.html>

Digital Natives and Specialised Digital Libraries: A Study of Europeana Users

Milena Dobрева¹, Emma McCulloch¹,
Duncan Birrell¹, Yurdagül Ünal², and Pierluigi Feliciati³

¹ Centre for Digital Library Research (CDLR), Information Resources Directorate (IRD),
University of Strathclyde, Livingstone Tower, 26 Richmond Street, Glasgow,
G1 1XH United Kingdom

{milena.dobрева, e.mcculloch,
duncan.birrell}@strath.ac.uk

² Department of Information Management, Hacettepe University,
06800 Beytepe, Ankara, Turkey
yurdagul@hacettepe.edu.tr

³ Department of Cultural Heritage, University of Macerata, via Brunforte,
63023 - Fermo, Italy
pierluigi.feliciati@unimc.it

Abstract. The alignment of user needs with the technical capabilities of modern digital libraries is an area attracting the interest of researchers and practitioners. Europeana, conceived with the intention of offering a single access point to European cultural heritage, has been developed in recent years with a continuous effort to identify and respond to the needs of a range of users. This paper presents a study of two user communities – young people and the general public. The study, conducted between October 2009 and January 2010, comprised a series of focus groups and media labs in Bulgaria, Italy, the Netherlands and the UK. A distinctive aspect of this study is that it combines questionnaire-based and verbal feedback gathered from users with evidence of user actions whilst undertaking a well-defined task. The paper presents the context and the methodology of the study, and some of the data gathered within the study which helps to understand better the attitude of digital natives towards specialised digital libraries. The data analysis supports several conclusions: specialised digital libraries require strong advocacy to target the “digital natives”¹ generation which tends to prefer general purpose search engines to specialised resources; young users are confident that they know how to use advanced search yet there is little evidence of their applying these skills in contrast to general public users; the perception of digital libraries differs in groups from different countries. The study contributes to the better understanding of some behavioural characteristics of users of digital libraries.

Keywords: Digital libraries, Europeana, user groups, user characteristics, user study methodology.

¹ The expression was introduced by Mark Prensky, see [1].

1 Background and Purpose

The modern digital libraries have to address multiple diverging requirements and expectations of the users and at the same time accommodate the rapidly changing technological novelties. Currently it seems that digital libraries are built basically to fill in specific gaps of provision and/or to find out how a specific technological solution could improve the functionality of a digital library. As Khoo, Buchanan and Cunningham noted

In the case of digital library researchers, the focus of research is often on technical issues (e.g., information retrieval methods, software architecture, etc.) rather than on user-centered issues. When these researchers turn to user based evaluations, they therefore often lack the necessary expertise to develop robust Human Computer Interaction (HCI) experiments, and their goals are typically limited to "proof of concept" tests, rather than prescribing user motivations or cognitive impacts. [2]

User needs are always mentioned amongst the priorities in the multitude of policy documents which shape the scope and characteristics of digitisation and accessibility online. However, in reality, users seem not to be consulted in detail. For example, in a review of the use of records in archival collections, A. Sundqvist wrote that “the general knowledge of user behaviour is a mixture of common sense, presumptions and prejudices” [3]. The findings of the Institute of Museum and Library Services (IMLS) reported that “The most frequently-used needs assessment methods do not directly involve the users” [4].

One possible explanation of the scarcity of user studies in the digital libraries domain is that many digital libraries exist to address “content gaps” and are built under the assumption that the user needs of professionals will cover the range of uses of the digital library. However, in the current setting when digital libraries also are used in a multicultural environment (which also includes the issue of multilinguality) user studies are needed to tailor digital libraries better to different user communities and to the individual users.

The currently existing models of digital libraries address users. For example, Users is one of the six domains within the DELOS DLRM (Digital Library Reference Model) alongside Content, Functionality, Quality, Policy and Architecture [5]. The 5S model introduces the basic notions of Streams, Structures, Spaces, Scenarios and Societies where users are part of the Societies [6].

Gathering evidence based, qualitative insights into the information behaviour of contemporary users would have very real impact on the future of *interface functionality*, *digital library policy*, *data quality* and possibly even on the *architecture* of digital libraries.

Recent developments in digital library design concentrate effort on the use of innovative search and browse tools, streamlined techniques for navigation and display, and the provision of personalised areas for search management and information sharing; such developments, however, remain unaligned to any thorough understanding of exactly *how* end users make use of such advanced tools, deploy new functionalities as part of search strategies and select resources for use – in other words how user behaviour alters depending on the scenario of use. Recently, an impressive range of

projects undertook evaluations of digital libraries and digitisation projects paying attention to various user-related issues:

- A *high-level formative evaluation* of the JISC Digitisation Programme Phase 2 [7] in the UK was made, addressing five key areas: *content creation; adoption of standards; innovation in sustainability and business models; resource discovery, personalisation and contextualisation; capacity building*.
- The academic impact of five projects undertaken in phase one has been evaluated within the University of Oxford's Survey on the Use of Digitised Resources [8], which released a *toolkit for the impact of digitised scholarly resources* (TIDSR).
- The Intute/JISC Digitisation Dissemination Project (IJDDiP) [9] identified *common research themes across digitisation projects*; addressed the issues of *reusability of online materials* and initiated scholarly discussion on *new collections and their impact on research*.
- The issues of *digitisation, curation and models for community engagement and involvement of the wider public* were also analysed in depth [10].
- The DiSCmap project [11] studied in detail the user needs for digitisation of special collections and produced a user-driven framework of priorities for digitisation alongside a list of collections nominated for digitisation.

Other recent studies such as the “Sustainability and Revenue Models for Online Academic Resources” [12], commissioned by the Strategic Content Alliance (SCA) in the UK, suggested business models to be used in the creation of online academic resources designed to take ephemeral ideas like “value”, “impact” and “usage” and apply them in practice.

Amongst the range of issues related to the benefits of digitisation, a clear and detailed understanding *of the ways in which the users interact with digitised content* is of central importance when considering issues relating to digital libraries. For example the LAIRAH Project [13] applied *log analysis* with the primary aim being to study factors contributing to the success and popularity of digital humanities resources. The impact of digitised collections on research and teaching was recently presented by OCLC [14], and a comprehensive overview the web interaction methods addressing user needs in the cultural domain was produced by the MINERVA project [15]. The TIDSR project synthesised quantitative and qualitative metrics for the impact of digitised resources.

These studies address different aspects of user interaction with the digital library. They can be positioned within the wider digital environment using the Information Triptych Model [16] which seeks to make clear the connection between User, Content and System. Fig. 1 contextualizes these areas and the evaluation metrics which could be used to assess the quality of digital libraries. Although multiple recent studies address the users, the information behaviour of different types of users as introduced by Wilson [17] still needs to be studied in more detail.

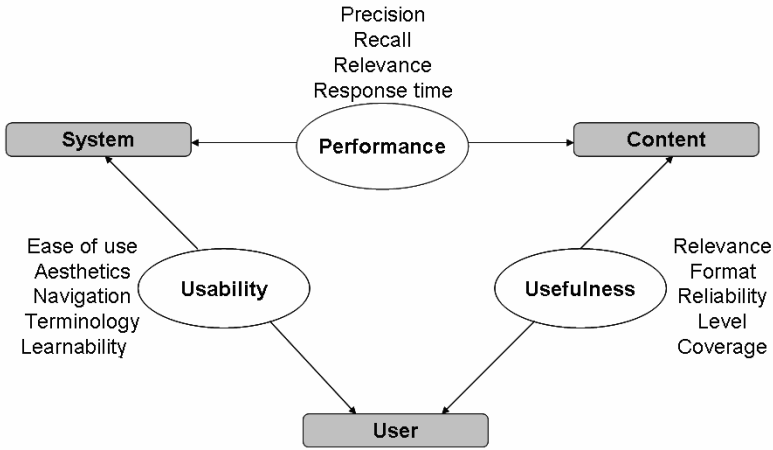


Fig. 1. Metrics related to usability, usefulness and performance mapped to the interaction triptych model

This paper presents a user study on one specific digital library, Europeana, and looks in more detail at the behaviour of young users.

1.1 Europeana and User Studies

Europeana² is a single access point for digitised cultural heritage materials provided by various European libraries, museums, archives, galleries, audiovisual collections and other memory institutions. This specialised digital library was launched by the President of the European Commission (EC) in November 2008. Currently it provides access to over 4.6 million objects with the aim of reaching a target of 10 million objects in 2010; more than 1000 institutions are providers of the cultural content in Europeana and their number and geographic coverage are steadily growing.

Europeana brings together an unprecedented amount of materials and thus plays the role of the major specialised aggregator in the cultural heritage domain for Europe. Europeana is able to offer a considerable range of content and is currently expanding through contributions made by projects from the Europeana group,³ supported by the EC and involving local and regional institutions and through the aggregation of content of various types (from libraries, archives, travel, cinema and TV, musical instruments, etc.). Europeana is more than an aggregator, however, since it also seeks to provide innovative ways of searching and visualising the rich cultural contents. This is being achieved through the gradual development of new versions of the interface with improved data organisation, search and browsing functionalities (the next one will be Europeana Rhine [19] which will be released later in 2010).

² <http://www.europeana.eu/portal/>; see also [18].

³ These projects are presented on <http://group.europeana.eu/>

Stakeholders developing Europeana are in regular discussion regarding how best to approach and serve its users. In fact, the concept itself of a “portal” includes the mission of “service provider”, an added value with respect to single sites. This extra sets aside the identity of the individual cultural subjects providing their data and deals directly with customer satisfaction.⁴ User studies for Europeana involve a combination of methods including gathering expert opinion, organising focus groups, carrying out observations of users and conducting a web survey. In the future Europeana will enrich knowledge of its users through log analysis; it also plans to work on the development of formal models of users (personae).

This paper presents some of the outcomes from a Europeana User and Functionality Study [21], which was coordinated by the Centre for Digital Library Research⁵ at the University of Strathclyde in Glasgow and implemented jointly with the University of Macerata,⁶ Italy, and the Emotion Lab⁷ of Glasgow Caledonian University.

1.2 The Europeana User and Functionality Study in the Context of Previous Studies of Europeana

This study aimed to address two specific user communities (young people and members of the general public) across four countries through a series of focus groups and media labs. Its purpose was to establish a better understanding of users’ expectations as well as the difficulties and stumbling blocks encountered while using the Europeana prototype.

The choice of countries in which focus groups and media labs were organised was informed by the outcomes of the preceding web survey of Europeana [22]. The Netherlands, the UK, Italy and Bulgaria were selected because they had differing response rates to the web survey (low, medium and high) and also different levels of contribution to Europeana in terms of resource provision. Their being geographically distinct and having different educational systems, especially in the area of information and computer literacy, also serves to create diverse conditions for the study. The school curricula in information technology differ both in terms of content covered and in terms of flexibility; information technologies is a mandatory subject in Bulgarian and Italian secondary schools. Roughly sixty percent of all secondary schools in The Netherlands now offer informatics as part of their curriculum [23]; in the UK it is optional; the exact coverage of the subject in Scotland is not specified, although the skills that need to be developed are clearly stated (see 5 to 14 Curriculum⁸). This could provide a basis to analyse whether the skills of young users as well as their

⁴ “For the users, a portal is surely only useful if it meets a real need that users have, and in a way with which they are comfortable. As such, the portal needs to do more than any of the current offers being presented. To facilitate this, there is need for continued work on ensuring interoperability of systems” [20].

⁵ <http://www.strath.ac.uk/cdlr/>

⁶ <http://www.unimc.it/>

⁷ <http://www.gcal.ac.uk/creates/centres/emotionlab.html>

⁸ <http://www.ltscotland.org.uk/5to14/resources/index.asp>

expectations differ substantially between countries and how educational background may be influencing them.

The study involved approximately the same number of participants in each of the different countries. Groups with secondary school children were studied in Sofia, Bulgaria and in Amsterdam, The Netherlands. University students were targeted in Fermo, Italy and members of the general public were recruited for the groups in Glasgow, Scotland. All group sessions were held between October 2009 and 15 December 2009.

The involvement of two types of user in the study – young people and members of the general public - was purposeful to provide a basis for analysing whether there are substantial differences within, and between, the user types. Unlike other European studies, this study:

- addressed participants' responses combined with analysis of the evidence of user actions during the completion of a standardised task in all the groups;
- included homogeneous groups and an equal distribution of the number of participants in each of the four countries.

The number of participants (total 89) is not sufficient to come to any statistically significant conclusions but the opinions and observations gathered are of interest and can be compared with the outcomes of other current user studies.

2 Study Methodology

All focus groups and media labs followed the same protocol:

1. Introduction to the study by the group moderator.
2. Completion of a pre-questionnaire (providing basic demographic information, indication on familiarity with Europeana, online search experience and cultural attitudes).
3. A concise introduction to Europeana, provided by the group moderator.
4. Discussion 1 (to gather first impressions of Europeana following a brief look at the site and its key features).
5. Completion of questionnaire 2 (to provide written feedback of first impressions).
6. Assignment (compiling a PowerPoint presentation about the local city).
7. Discussion 2 (to gather deeper impressions of Europeana following approximately 30 minutes' interaction with the resource).
8. Completion of questionnaire 3 (to provide written feedback on deeper impressions of Europeana).
9. Conclusion of study, by moderator.

In the case of media labs, discussions took the form of a conversation between the moderator and the participant. In the case of the focus groups, these were common discussions with all group members facilitated by the moderator.

The assignment was designed to incorporate eight different usage scenarios designed to encourage the participants to search for different types of objects on a predefined topic. For all groups the general assignment was to build a virtual portrait of their city. More specifically, the users were required to search for texts (hinted at

by a slide entitled “What people write about Sofia/Rome/Amsterdam/Glasgow); images; audio and/or video; materials on the same object from different times; materials on a very specific predefined subject (like a landmark or an event or a person – e.g. Glasgow School of Art); a particular historical event where primary resources were expected to be gathered; materials of personal choice related to the presentation theme; and finally, identifying the providers of digital objects who contributed the highest number of objects on a particular topic, identifying what was found to be most useful about Europeana and suggesting areas in which material may be lacking – which encouraged reflective practice by the students. This range of scenarios required users to formulate searches that target a range of metadata fields to retrieve various types of materials; some searches would be more efficient if the users employed advanced search options. This approach facilitated an assessment of which usage scenarios are easy to satisfy and the stumbling blocks that users of the Europeana prototype may encounter.

A distinguishing feature of this study is that it combines feedback gathered from users with evidence for their behaviour and reported views. Contributions to discussions were supplemented by responses to questionnaires and further consolidated by their search strategies and subsequent selection of materials held within Europeana (which were able to be assessed by accessing participants’ MyEuropeana results) and by examining the content transferred from Europeana to their PowerPoint presentations. The protocol was designed so that feedback gathered from the users at various stages of the study effectively reflected their *first impressions* and expectations (following a brief presentation providing an overview of Europeana and its key features) before the actual assignment; *deeper impressions* (after the users worked on the assignment) which help to ascertain whether or not the nature of the service and its delivery met the expectations expressed earlier, and *lasting impressions*, showing the intentions to use Europeana in the future, following completion of the assignment and participation in a subsequent group discussion (or individual discussion, as in the case of the media labs). The series of media labs conducted in Glasgow provided an additional means of feedback due to the collection of physiological data. Facilities enabled eye tracking data to be gathered, pinpointing the gaze of participants throughout the assignment, as well as the duration of their focus on any one area of the screen/interface.

The study aimed to gather evidence and observations on the information behaviour of the participants and to see if there are substantial differences across countries; it did not formulate any particular hypotheses in advance. Although interest in the “digital natives” is considerable, little research has been undertaken to address the existing gap in knowledge between understanding digital natives’ behaviour and the development of digital libraries to provide the functionality and usability needed to satisfy these behaviours [24].

3 Demographic Characteristics of the Participants

The participants were resident in four countries. The total number of participants was 22 in Bulgaria, 20 in Italy, 23 in the Netherlands and 24 in the UK. About half of the

participants were aged between 15 and 18 (46%); almost one third were 19-24 (28%). 44% of all participants were male and 56% were female. More than three quarters of the participants were students (79%), the rest were members of the general public with an interest in art and culture.

Most of the participants (98%) had not used Europeana before although some had seen it. It was not widely known to the participants, with 91% reporting that they were not familiar with the Europeana logo.

Most of the participants (97%) reported that they frequently search for materials online and 67% of the participants claimed to use advanced online search features. Search by phrase was reported as being used most frequently within the groups of young users, while search by date and using Boolean operators were more popular amongst the general public users from the UK.

The most searched file types by participants are text (45%) and images (22%), with fewer reporting searching for video clips and audio files (about 17%). Table 1 reports on considerably different proportions in the preferred objects which the participants search for most frequently in the different countries.

Table 1. Objects/file-types searched for on a weekly basis by participants

Object	Country									
	Bulgaria		Italy		The Netherlands		UK		Total	
	N	%	N	%	N	%	N	%	N	%
Texts	14	26.9	18	60.0	22	46.8	22	53.7	76	44.7
Images	15	28.8	3	10.0	9	19.1	10	24.4	37	21.8
Video clips	11	21.2	3	10.0	10	21.3	5	12.2	29	17.1
Audio files	12	23.1	6	20.0	6	12.8	4	9.8	28	16.5
Total	52	100.0	30	100.0	47	100.0	41	100.0	170	100.0

4 Findings

4.1 Specialised Digital Libraries Need Strong Advocacy to Target “Digital Natives”

The study aimed to ascertain the extent to which a specialised digital library like Europeana, providing access to shared European cultural heritage materials, is attractive to young users. The very idea of using specialised websites was not very popular with the young participants in the study. Asked about the type of websites they would use to search for a specific type of digital resources (e.g. images), the young users in all countries showed clear preference to general search engines (91.3% in Bulgaria, 90.0% in Italy, 78.6% in the Netherlands) and not to specialised websites (e.g. Flickr⁹). The Bulgarian and Dutch groups consisted of secondary school students who

⁹ <http://www.flickr.com>

seemed to be most happy with the use of general purpose search engines even when searching for a specific type of resources. Younger users, in fact, came to Europeana with the expectation that, regardless of its domain or specialism, an online resource should incorporate the variety of interface features and Web 2.0 functionalities found on more general sites.

The general public users showed highest (25.5%) preference for the use of specialised web sites in such cases.

This observation based on the questionnaire responses of the participants was echoed by their responses to an open-ended question about their website preferences (see Figure 2).

All websites appearing in participants' responses are summarised in Table 2. Again, the responses of digital natives show clear preference for general purpose websites. Amongst the search engines, Google is the most popular across all groups (62%); Wikipedia is the second most popular (36%). In the UK the participants had other preferences which could be attributed to the background of the group i.e. members of the general public with cultural interests. In the discussions within the focus groups it was noticeable that there is a tendency to compare Europeana with Google or Wikipedia. This is understandable, and suggests that any specialised digital library would be more attractive to digital natives if it can clearly communicate its advantages.

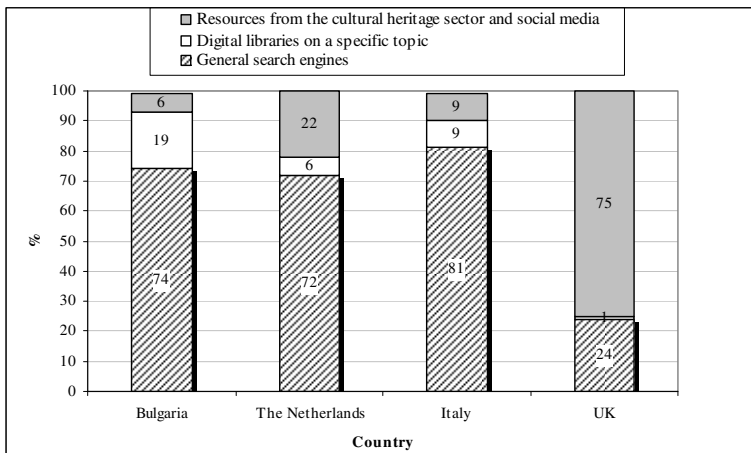


Fig. 2. Preferences for the use of different types of resources

Another indication of the difficulty in introducing specialised digital libraries to digital natives was the response to the question asking whether or not the participants would use Europeana in the future. Here the groups showed significant differences. The most positive response was received in Italy where three quarters of the participants replied that they would use Europeana in the future; in Bulgaria half of the

Table 2. Data on preferred Websites for school/work use

General search engines	Bulgaria		The Netherlands		Italy		UK		Total	
	N	%	N	%	N	%	N	%	N	%
Google	16	70	21	57	13	50	15	79	65	62
Wikipedia	7	30	14	38	13	50	4	21	38	36
Yahoo	-	-	2	5	-	-	-	-	2	2
Total	23	100	37	100	26	100	19	100	105	100
Digital libraries on a specific topic	Bulgaria		The Netherlands		Italy		UK		Total	
	N	%	N	%	N	%	N	%	N	%
Pomagalo	4	67	-	-	-	-	-	-	4	31
Deviantart	2	33	-	-	-	-	-	-	2	15
Flickr	-	-	-	-	1	33	1	100	2	15
arte e motori	-	-	-	-	1	33	-	-	1	8
Photobucket	-	-	1	33	-	-	-	-	1	8
Encyclopedia drammatica	-	-	1	33	-	-	-	-	1	8
Youtube	-	-	1	33	-	-	-	-	1	8
MSN encarta	-	-	-	-	1	33	-	-	1	8
Total	6	100	3	100	3	100	1	100	13	100
Resources from the cultural heritage sector and social media	Bulgaria		The Netherlands		Italy		UK		Total	
	N	%	N	%	N	%	N	%	N	%
Provider for cultural sector	-	-	-	-	1	33	18	31	19	25
Online	-	-	1	9	-	-	11	19	12	16
Educational/Academic resource										
National / University Library Service	-	-	-	-	-	-	11	10	11	15
Broadcast and Print Media	-	-	-	-	-	-	6	10	6	8
online sources										
Social Networking sites	-	-	-	-	-	-	5	8	5	7
OPAC	-	-	-	-	1	33	-	-	1	1
Other	2	100	10	91	1	33	8	14	21	28
Total	2	100	11	100	3	100	59	100	75	100

participants claimed to be happy to use Europeana again. Responses in Amsterdam were mostly negative with 30 statements about what they do not like about Europeana (ranging from criticisms on its content to critiques on the functionality); in the UK there were 19 negative and 12 positive statements.

The general preference of young people, even when required to search for a specific type of resource, is to use a general search engine. Hence the strategy to introduce the use of a specialised digital library like Europeana needs to be very well-targeted and compelling. *Such specialised digital libraries should have very strong advocacy policies which should clearly show the resource's advantages. Simply building a good and trustworthy resource is not enough to attract the young users.*

4.2 Information Literacy: Skills of Young Users Can Be Improved

Another area of the study concerned existing online search competence of users - an important aspect of their information literacy skills. 72% of participants in Bulgaria, together with 60% in Italy, 57% in the Netherlands and 75% in the UK reported that

they frequently use advanced online search features and in general, young users seem reasonably confident about their online search skills. Advanced search is used very often or often by 77.3% of the respondents in Bulgaria, 60% in Italy, 56.6% in the Netherlands and 75% in the UK. Here the levels of confidence are slightly different but still relatively high for all of the participating countries.

Relating to the use of advanced search features, and which of these features users feel confident with, young users' responses indicated a strong preference for *searching by phrase* (90.9% in Bulgaria, 87.5% in Italy, 81.8% in the Netherlands and 43.8% in the UK). *Searching by date* was the second most popular type of search reportedly used and the reported use of *Boolean operators* was low amongst young users (0% in Bulgaria, 4% in Italy, 13% in the Netherlands) compared to 63% in the general public group in the UK. This suggests that general public and older people are really using advanced search features while young people seem confident in their use but there is a lack of correspondence with the levels of response on the use of Boolean operators.

We compared the initial data collected using questionnaires with the queries saved by participants during the assignment (all participants registered in MyEuropeana and were asked to save their searches). Analysis of the queries recorded by participants shows that although their level of confidence in the use of advanced search was high, the actual searches run by young users were ONLY simple searches. A total of 17 Boolean searches were conducted by members of the general public and not by the young users. Examples include: 'Pere Lachaise OR La Chaise'; 'Glasgow AND school AND art'; 'Glasgow AND South Africa'; 'Glasgow AND George Square AND 191*'.

A summary of the types of searches saved during the focus groups is presented in Table 3. Most popular amongst digital natives were the searches on proper nouns, corresponding to the nature of the task.

From the evidence of user searches it can be concluded that although young users are confident in their online search skills, the actual use of a well developed resource like Europeana did not illustrate efficient use of the extent of search functionality provided; in contrast, the group of general users did employ the advanced search options. *This probably means that the information literacy of the young generation can be improved upon.*

4.3 Perceptions of Digital Libraries Differ in Different Environments

The study included participants from four countries and although the number of participants is not sufficient for statistically significant results, the study provides some indication of areas where differences between user groups exist and which may be of interest to study further. One such area is the perception and expectations of Europeana, found to vary between the participants from the Netherlands, Italy and UK, shown when asked to express their initial impressions and ideas about the website after being familiarised with it.

Table 3. Types of searches

Linguistic breakdown	Sofia FG		Amsterdam FG		Glasgow FG		Glasgow ML		Total	
	N	%	N	%	N	%	N	%	N	%
Proper nouns Examples: Sofia; Royal Palace; Glasgow School of Art; Buchanan Street	41	85	32	53	67	64	63	65	203	65
Nouns Examples: Palace; Marijuana	2	4	1	2	0	0	0	0	3	1
Two-word compound nouns Examples: Sofia culture; coffee shop; art nouveau; Glasgow fashion	2	4	17	28	22	21	18	19	59	19
Phrases Examples: Amsterdam civilian perspective; Glasgow city of culture; Sounds from Sophia	3	6	10	17	15	14	18	19	46	15
Total	48	100	60	100	104	100	99	100	311	100

In Bulgaria, initial impressions were captured by asking participants to select the most relevant description of Europeana from a list of dichotomic pairs (words with opposite meanings, e.g. “dull” – “exciting”). The most popular choice in Bulgaria was “easy to use” (19 of 22 participants), and one quarter of the participants also chose the terms “unique”, “attractive”, and “exciting”. Each of the following descriptions was chosen once: “similar to other sites”, “fun”, “badly organized”.

Fig. 3 summarizes the overall estimates given by participants for all dichotomic pairs by country. The higher the estimation is, the more positive (the “negative” word was at the low end of the scale – 1, and the “positive” was at the highest value of 10). The results show differences in the perception of Europeana: generally the UK participants were more critical in their initial reaction to it while Italian participants seemed to be more excited about Europeana.

The reasons why some groups are more critical than others requires further study (one possible explanation could be that more experienced users are more critical, but this seems simplistic and is not supported by enough evidence). Further research on the country-specific perception of digital libraries would help to fine-tune *personalisation of digital library interfaces, not only to the personal characteristics of the users but also to their cultural background.*

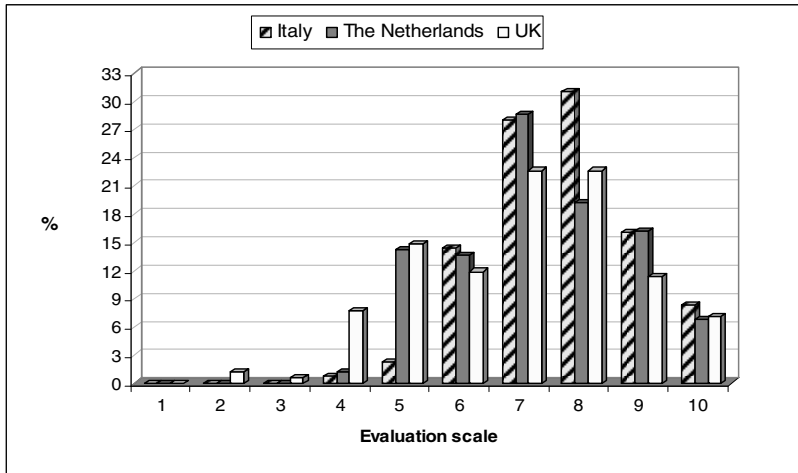


Fig. 3. Summary of responses on dichotomic pairs

5 Discussion

The understanding of the digital libraries' users and their information behaviour is far from simple. While positioned as a priority, users of cultural heritage materials have not been consulted in detail and their information behaviour is not well studied. The digital library community should be better informed about the mechanisms capable of increasing interest in such specialised resources for digital natives. Although a very detailed understanding on the specific characteristics of digital natives does not yet exist, it seems that we know about typical information behaviour and suppose that it is homogeneous, mostly in terms of information skills – but there is no extensive evidence to back up such beliefs.

The study presented in this paper involved a relatively small number of participants and yet one of its interesting features was the combination of methods where the self-evaluation of participants could be compared to their actual information seeking behaviour. A popular recent study based on log analysis, CIBER's "Google Generation" [25] started to build a portrait of contemporary internet users. It showed that information seeking is horizontal ("people view just one or two pages from an academic site and then 'bounce' out, perhaps never to return"; users spend most of their time on navigation and not on the actual use of resources; the users act as viewers – spending time which is obviously not sufficient to read the documents they are viewing; and they develop "squirreling behaviour" – downloading multiple documents (it is difficult to study how many of those are used afterwards and how).

We could add several characteristics to this profile of digital natives in the context of specialised electronic library usage:

1. Preference for general search engines. Digital natives prefer to use general search engines such as Google and Wikipedia rather than specialised resources.
2. High search confidence is not necessarily backed up with skills. Young users are confident in their skills for online searching, but in carrying out tasks which hinted at using advanced searches within a digital library which supports them, not one participant amongst the young users taking part in the study ran an advanced search. This could suggest that young users need to undertake more practical tasks, enabling them to practice their online searching skills.
3. Users in different countries perceive digital libraries in different ways. Some users are more critical than others. This is definitely an area where more research is needed. The differences in the education in information technologies in the participating countries does not seem to correlate with the different perceptions. Variance could be attributed to individual differences, rather than to geographical diversity, but further investigation is required.

One of the strengths of this study is that data supplied by the participants through questionnaires and discussions is backed up with an analysis of the evidence on user actions, such as queries, populated presentations and eye tracking data. Although user study methods are generally well-established and described [26], an area where more guidance is needed is how to combine different user study methods efficiently. The principle of evidence-based research was followed in the present study, which helped to identify some differences between the self-estimation of the participants and their actual behaviour. Such studies provide valuable findings which can be integrated with the recommendations in existing handbooks on cultural web user interaction (see [27]).

6 Conclusions

The methodological approach in this study addresses a clear objective and involves well-defined user groups.

The alignment of user needs with the technical and political capabilities of the modern digital library is a complicated and usually expensive task. Specific user needs should be studied in relation to particular foci, including (1) ease of use and intuitiveness of the digital library; (2) identification of 'future' user needs as the young generation grows up; (3) styles of use of the Europeana prototype for knowledge discovery amongst young users; (4) expectations and trustworthiness; (5) similarities and differences in the groups from different countries; (6) possible recommendations for prototype development in line with user needs.

As a matter of priority, the needs of disparate, yet specific, user groups need to be studied in depth. The digital natives' generation is of key interest because it has current expectations and also typifies the nature of a wide range of future users (general as well as professional) of Web content.

The study involved two user groups, young people and members of the general public with a clear interest in the culture/arts domain. This facilitated a comparison of user characteristics, and although the study involved a relatively small number of participants, it provided helpful feedback for Europeana and raised some significant questions for the continued study of the digital natives' information behaviour.

Acknowledgments. The authors would like to acknowledge the support of the EDL foundation; Jill Cousins, Adeline van den Berg, Anne Marie van Gerwen and Milena Popova from the Europeana office. The discussions with Prof. Ian Ruthven (University of Strathclyde) and Prof. Jonathan Sykes (Glasgow Caledonian University) were particularly helpful for the study. We would also like to thank all the participants in the study whose commitment and insights were extremely valuable.

References

1. Prensky, M.: Digital Natives, Digital Immigrants. On the Horizon 9, 1–6 (2001), <http://www.marcprensky.com/writing/Prensky%20-%20Digital%20Natives,%20Digital%20Immigrants%20-%20Part1.pdf>
2. Khoo, M., Buchanan, G., Cunningham, S.J.: Lightweight User-friendly Evaluation Knowledge for Digital Libraries, D-Lib Magazine, 15 (2009), <http://www.dlib.org/dlib/july09/khoo/07khoo.html>
3. Sundqvist, A.: The Use of Records – A Literature Review. Archives & Social Studies: A Journal of Interdisciplinary Research 1, 623–653 (2007)
4. IMLS: Assessment of End-User Needs in IMLS-Funded Digitization Projects, <http://www.imls.gov/pdf/userneedsassessment.pdf>
5. Candela, L., et al.: The DELOS Digital Library Reference Model - Foundations for Digital Libraries. Version 0.98, http://www.delos.info/files/pdf/ReferenceModel/DELOS_DLReferenceModel_0.98.pdf
6. Goncalves, M., Fox, E., Watson, L., Kipp, N.: Streams, Structures, Spaces, Scenarios, Societies (5s): A Formal Model for Digital Libraries. ACM Transactions on Information Systems 22, 270–312 (2004)
7. Formative Evaluation of the JISC Digitisation Programme Phase 2, http://www.jisc.ac.uk/media/documents/programmes/digitisation/digevalfinalreportf2_final_002.pdf
8. Meyer, E.T., Eccles, K., Thelwall, M., Madsen, C.: Usage and Impact Study of JISC-funded Phase 1 Digitisation Projects & the Toolkit for the Impact of Digitised Scholarly Resources (TIDSR), <http://microsites.oii.ox.ac.uk/tidsr/>
9. Intute / Jisc Digitisation Dissemination Project, <http://www.jisc.ac.uk/media/documents/programmes/digitisation/intutefinal.pdf>
10. Digitisation, Curation and Two-Way Engagement, http://www.jisc.ac.uk/media/documents/programmes/digitisation/dcatwefinalreport_final.pdf
11. DiSMap: Digitisation in Special Collections: Mapping, Assessment, Prioritisation, <http://www.jisc.ac.uk/whatwedo/programmes/digitisation/reports/discmap.aspx>

12. Guthrie, K., Griffiths, R., Maron, N.: Sustainability and Revenue Models for Online Academic Resources, http://sca.jiscinvolve.org/files/2008/06/sca_ithaka_sustainability_report-final.pdf
13. Warwick, C., et al.: The LAIRAH Project: Log Analysis of Digital Resources in the Arts and Humanities, <http://www.ucl.ac.uk/infostudies/claire-warwick/publications/LAIRAHreport.pdf>
14. Proffitt, M., Schaffner, J.: The Impact of Digitizing Special Collections on Teaching and Scholarship: Reflections on a Symposium about Digitization and the Humanities, <http://www.oclc.org/research/publications/library/2008/2008-04.pdf>
15. Handbook on Cultural Web User Interaction, <http://www.minervaeurope.org/publications/Handbookwebuserinteraction.pdf>
16. Fuhr, N., et al.: Evaluation of Digital Libraries. *International Journal on Digital Libraries* 8, 21–38 (2007)
17. Wilson, T.D.: Models in Information Behaviour Research. *Journal of Documentation* 55, 249–270 (1999)
18. Europeana i2010: Europeana – A Single Access Point to Europe’s Cultural Heritage. Europe’s Information Society Thematic Portal, http://ec.europa.eu/information_society/activities/digital_libraries/europeana/index_en.htm
19. Bloomberg, R., et al.: Functional Specification for Europeana Rhine Release, D3.1 of Europeana v1.0 project (public deliverable) (2009)
20. Miller, P.: The Concept of the Portal. *Ariadne* 30 (2001), <http://www.ariadne.ac.uk/issue30/portal/>
21. Dobрева, M., et al.: User and Functional Testing, http://version1.europeana.eu/c/document_library/get_file?uuid=1c25ae28-9457-4b0f-be62-654a7cf6c5b7&groupId=10602
22. Europeana – Online Visitor Survey. Research Report, IRN Research, http://www.edlfoundation.eu/c/document_library/get_file?uuid=e165f7f8-981a-436b-8179-d27ec952b8aa&groupId=10602
23. Leeuw De, K., Ponse, A.: Informatics and Secondary Education in the Netherlands (2008), <http://staff.science.uva.nl/~alban/publist/ISEN.pdf>
24. Tonta, Y.: Dijital Yerliler, Sosyal Ağlar ve Kütüphanelerin Geleceği (Digital Natives, Social Networks and the Future of Libraries). *Türk Kütüphaneciliği* 23, 742–768 (2009)
25. CIBER ‘Google Generation’ Project. Information Behaviour of the Researcher of the Future, <http://www.ucl.ac.uk/infostudies/research/ciber/downloads/>
26. Bryan-Kinns, N., Blandford, A.: A Survey of User Studies for Digital Libraries, RIDL Working Paper, London (2000)
27. Feliciati, P., Natale, M.T.: Handbook on Cultural Web User Interaction, First edition, MINERVA eC WG5, Rome, <http://www.minervaeurope.org/publications/handbookwebusers.htm>

Impact of Recent Trends in Information and Communication Technology on the Validity of the Construct Information Literacy in Higher Education

A.A.J. (Jos) van Helvoort

The Hague University of Applied Sciences, Academy of ICT&Media, Post Box 13336,
2501 EH The Hague, The Netherlands
a.a.j.vanhelvoort@hhs.nl

Abstract. The objective of this paper is a reflective discussion on the validity of the construct Information Literacy in the perspective of changing information and communication technologies. The research question that will be answered is: what is the impact of technological developments on the relevance of the Information Literacy concept? Technological developments that will be discussed are:

- content integration (federated search engines)
- amateur publishing (user generated content)
- use of social networks to find information
- personalisation and push technology
- loss of context / fragmentation of information.

Research methods: desk research and critical analysis of the results that were found. The analysis of the influence of the discussed technologies on the Information Literacy concept is represented by arrow diagrams.

Findings: The Information Literacy concept refers to a set of sub skills varying from retrieval skills to critical use of scholarly information. Changing technologies reduce the significance of the more instrumental sub skills of the Information Literacy concept. On the other hand, higher order cognitive skills (for instance critical evaluation of resources and analysis of content) become more and more important for students and professionals who try to solve their information problems.

The paper concludes with a description of the facets of the Information Literacy concept that need extra attention in the education of the knowledge workers of the future.

Keywords: Information technology, communication technology, information literacy, higher education.

1 Background and Purpose of the Paper

The title of this contribution suggests a problem with the appropriateness of the paradigm Information Literacy (IL), due to developments in information and communication technology. The assumption is that the importance of information literacy education is not recognised by many of our colleagues, who teach in faculties and disciplines

different from the Information Studies Departments. Information Literacy as an important learning goal is mainly defended by librarians and by some educational researchers but is certainly not accepted by all higher education teachers and their students. In a paper in the *Journal of eLiteracy* 2005 Peter Godwin [1] gives a couple of reasons for this “academic apathy and disinterest” in Information Literacy:

- The confusion between information literacy and internet skills: academics see that most of their students are very skilful in internet searching and for that reason they conclude that the students are competent information problem solvers. This belief that good internet skills result in good information problem solving skills is very ingrained and hard to change.
- Some academics are very good information finders, using their peers in their social networks and recommendations they receive from alerting services and the scholarly journals they read. These information sources combined with the use of easy to handle general search engines are “good enough” for them. They do not realise the complexity of information selection for novices in their discipline.
- Academics are often more interested in teaching their subjects than in training general skills. They defend attention for their discipline based issues in the curricula and do not want them to become overwhelmed by more general items such as writing skills and information searching.
- Finally, some academics are frustrated by the pace of change of the information and technology tools and the lack of standardisation (different interfaces, passwords etc). “Why can’t it be as simple as the Google search box?” they ask themselves.

From the Information Studies viewpoint, I do not wish to accept all the objections of these academic colleagues. In this contribution I want to share with you my arguments as to why Information Literacy is still (and in my opinion particularly these days) an appropriate learning goal for Higher Education. This will be done by reviewing some developments in Information and Communication Technology (more specifically trends that are well known as “Web 2.0”) and by answering the main research question: what is the impact of those technological developments on the relevance of the Information Literacy concept? First of all, however, we will elaborate on the content of the construct Information Literacy.

1.1 Existing Conceptions of Information Literacy

Information Literacy has been discussed for more than thirty years in international Information Science literature. Most authors on Information Literacy agree that the concept refers to a person’s broad ability to use information from external resources, including a set of sub skills such as the abilities to:

- recognise an information need
- formulate a search question
- choose relevant information sources
- use ICT to consult information sources
- select, evaluate and organise the information that is found
- (re)use and disseminate the information [2], [3].

Although most of the authors do not really differ in their definitions for the Information Literacy concept, phenomenographic research¹ makes clear that in the context of education at least two different conceptions of the phenomenon can be distinguished:

- Information Literacy as a set of skills to solve specific information problems. In this conception Information Literacy is school task oriented and helps students to complete the assignments that they receive from their teachers. This conception of Information Literacy is called by Bruce the “information process conception” [5], [6] but her “information technology conception” and “information sources conception” are included in the educational models for this school task oriented conception of Information Literacy. Berkowitz & Eisenberg’s Big6-model [7] is a well known example of the task oriented conception of Information Literacy.
- Information Literacy as a personal knowledge base. In this conception Information Literacy is no longer associated with a specific task but is experienced as an individual’s ability to gather information and to organise it in a personal mindset for use at an appropriate time [8]. In the personal knowledge base conception of Information Literacy the sub skills are used by a person in a strategic way to prepare him/herself for information use in the long term. It includes the construction of mental models (“know what”), cognitive strategies (“know how”), cognitive tasks (“know when”) and self knowledge.

The six sub skills listed above play a role in both conceptions of Information Literacy.

2 Current Relevance of the Information Literacy Concept

Although the concept of Information Literacy in itself is not related to the digital manifestation of information, there is no doubt that the attention given to IL in the eighties of the last century was a result of the growing use of computers, computer networks and the availability of new media to store and to retrieve large amounts of information. It is of course not coincidental that the most authoritative publications such as the ACRL standards and Christine Bruce’s book *The Seven Faces of Information Literacy* were published in the years that the Internet developed into the World Wide Web, which created easy access to digital information for all citizens of the developed world.

However, because ICT has not stood still since the year 2000, it might be that technological developments have made the Information Literacy concept less appropriate for higher education these days or, on the contrary, even more important than ever before. This observation was the motivation to start the research that is described in this contribution.

¹ Phenomenography is a qualitative research method that is aimed at “description, analysis, and understanding of experiences” [4]. Phenomenographic research of Information Literacy is introduced by Christine Bruce.

2.1 Research Methods

Literature on technological developments in Library and Information Science was retrieved and collected by monitoring the main Dutch library bloggers for more than five years. The Netherlands has an active community of “bibliobloggers” that daily refers to books, articles, reports etc. from all over the world. These references to international LIS literature were managed with the Refworks research management tool and were analysed to find the main trends. In the appendix you can find a selective bibliography with the titles on recent developments in information and communication technology that were used to write the overview in the next section of this paper.

3 Overview of Developments in Information and Communication Technology and Their Impact on the Information Literacy Concept

3.1 Easy Access to a Variety of Bibliographic Tools and to Full Text Documents

For many years library and information professionals have adopted the idea of the ‘invisible’ or the ‘deep’ web [9]. The statement was that the most trustworthy information was in databases that cannot be indexed by the general search engines. That was the reason why Information Literacy training often included the use of a lot of different bibliographic databases and a variety of user interfaces. Moreover, the information users had to learn how to find the full text documents of the titles they retrieved from abstract services like ERIC, LISA or MEDLINE. Not surprisingly, for the average information users it was not clear why the retrieval systems offered by the library were not as easy to use as the popular search engines Google and Yahoo. This was more than true from the moment that general search engines made it possible to present the results from dynamic web pages and the records of bibliographic databases [see also 10] and the launch of general search engines like Google Scholar and Scirus for the retrieval of academic information.

For that reason information service providers such as libraries and information centres have invested a lot of time and money in the development of easy to use Meta Search or Federated Search Engines, combined with Open URL services that link to the full text documents available in the digital library collection of the institution (“link resolvers”). A good example of this kind of integrated search is the Summon search engine that provides search facilities similar to Google but with well indexed academic content (see for instance the website of the University of Liverpool Library).

As expressed in Figure 1, the availability of such easy to use integrated search facilities makes the knowledge of resources and the availability of ICT skills less significant for the average information users. In general, integrated search engines do not provide the more advanced search facilities that are provided by the individual bibliographic databases but for most of the amateur searchers their results are “good enough” [11, 12]. This is at least true for solving information problems in the context of specific research or study tasks. For Information Literacy interpreted as a skill to be

kept informed of current developments in a discipline, knowledge of specific resources and the use of more advanced retrieval tools such as RSS aggregators continue to be important. Knowledge of specific resources can also be useful to evaluate the quality of the documents found on the results page. An article on the history of cancer research, for instance, is considered to be more reliable if Google has retrieved an abstract of the article from the PUBMED database.

recognise an information need		easy access to bibliographic tools and full text documents
formulate a search question		
choose relevant information sources	←	
use ICT to consult information sources	←	
select, evaluate and organise the information that is found		
(re)use and disseminate the information		

← reduced significance

Fig. 1. Impact of easy access to bibliographic tools and full text documents on the Information Literacy sub skills

3.2 Easy to Use Publishing Tools

The World Wide Web has developed into a platform upon which the publication model that has been used for many years has changed dramatically. The classical information chain with an author, a publisher, a mediator, such as a library, and the reader as a consumer, is no longer everyday practice. Content creators can put their articles, stories, pictures, podcasts and audiovisuals directly on the internet. They can even use publication channels like Blogger and YouTube at no cost. Criticasters such as Andrew Keen fulminate against this practice as a threat to the creative industry [13]. However, from the Information Literacy point of view, the rise of these possibilities for amateur publishing has significance for two other reasons.

On the one hand it increases the importance of the ability to evaluate critically the information resources on reliability, completeness and accuracy. The fact that information is found on the internet does not of course mean that it is worthless. Not only amateurs but also professional researchers and journalists use internet publication channels such as bulletin boards, weblogs, wiki's and forums to publish their preliminary research results or drafts of their papers and articles. All these publication types have their rules, formats and structures [12]. For internet users it becomes

important to develop their ability to recognise these genres and to use them to filter the more trustworthy information from the more doubtful. On the internet the qualifying role of publishers and journal editors is, indeed, often missed which means that selecting and evaluating the information that is found is harder than in the classical publishing situation.

On the other hand, the ease of online publishing creates a magnificent playground for students to train their research, debating and writing skills on a more informal platform, where they can nevertheless receive comments from their peers and their professors. It is, in other words, a great platform for their first steps in the academic publishing world [14]. Information Literacy should not be restricted to information retrieval but should also include the critical use of information in the students' own knowledge products. The use of publishing channels on the internet should be promoted for educational reasons but it also increases the significance of the students' skills in writing and publishing their products, which include their abilities to understand and to avoid plagiarism and to respect the intellectual properties of other content creators. It seems that handling these issues is rather simple but in practice students struggle a lot with them, just as with the avoidance of situations which can harm their privacy [15].

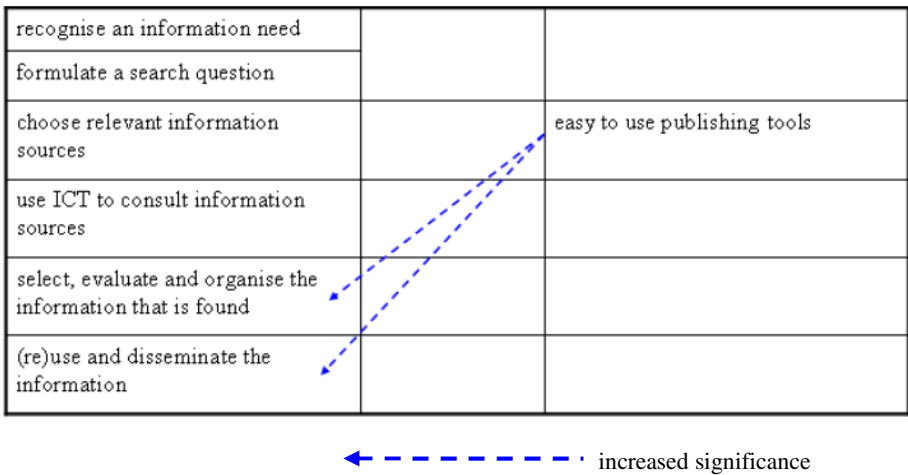


Fig. 2. Impact of easy to use publishing tools on the Information Literacy sub skills

3.3 Network Functions of the World Wide Web

Professional information seeking behaviour is not restricted to direct keyword searching but implies the reflection on relevant search results and further retrieval on footnotes, citations, supplied subject headings and key authors. The modern retrieval tools on the World Wide Web facilitate these berrypicking techniques [16] by presenting those elements as clickable hyperlinks. Starting from one article in a retrieval system,

the information searcher can, for instance, jump to the main subject heading in the description and retrieve with one click all other articles with the same subject heading. Another example of a retrieval tool that uses the networked character of the World Wide Web is the ‘Cited by’ button of Google Scholar.

These examples are from the field of scholarly publishing but they have their counterparts in the more popular places on the World Wide Web such as the social bookmarking site Delicious and the recommendations in online shops like Amazon. Hopping from one point on the World Wide Web to another is what Carroll [17] called a “meandering search strategy”. It reduces the importance of a good search question formulation but makes the searcher’s evaluation skills a much more critical factor, since this strategy (if it is based on user generated content) depends on the expertise of the people that have tagged the websites or have written the recommendations.

Another example of the networked character of the new World Wide Web is the social function. In the first decade of our new century the World Wide Web has developed from a place to find information to a platform primarily used for communication and social activities. Young people spend a lot of time on instant messaging platforms such as Windows Live Messenger (formerly MSN) and social networking sites like MySpace and Facebook. Just like professional researchers who consult their peer network for recommendations to solve their information problems, there are a lot of students who use their social network on the internet to get their information for completing a study task. As the most popular Dutch Library Blogger has recently expressed: “the new way of searching is asking questions” [18].

However, just as in the real world also on the internet one friend is not the same as the other, and the quality of the answers depends a lot on the expertise of the person that gives them. Also from this point of view, there is no doubt that the use of networks on the World Wide Web for information problem solving makes the role of evaluation skills increasingly important..

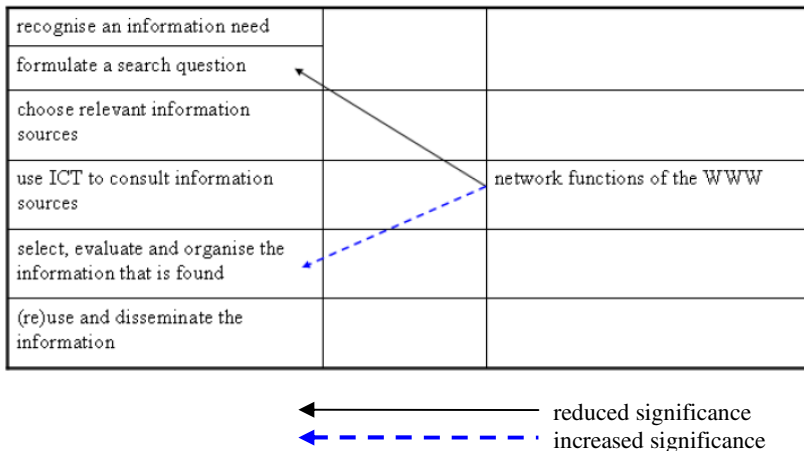


Fig. 3. Impact of the use of network functions of the WWW on the Information Literacy sub skills

3.4 Personalisation of the Information Dissemination

Professional information users not only search actively for information for solving a specific problem but also monitor resources to keep in touch with the newest publications and information on the World Wide Web. This more passive way of information seeking nowadays is strongly supported by push technologies that use personal profiles of the subscribers for a service. An interesting example of this current technology is customised traffic information based on the location of your smart phone with a GPS signal (Global Positioning System).

In the academic and professional world these kinds of personal alerting services were introduced in the 1960's (we know them as 'Selective Dissemination of Information' or 'Current Awareness Services') but the use was restricted to a small number of early adopters. Nowadays the push technology is used by a lot of people for subscription to electronic newsletters and discussion lists but professional information users also use it to be notified of new content in a database or on a website that matches a personalised query or 'profile'. The alerting is mostly done by sending an email and sometimes by text messaging (SMS). However, after some time a lot of people experience it as information overload or as spam. Also irritating are the difficulties to unsubscribe to the alert following an automatic subscription that is used by a commercial website.

An alternative to email alerting is the use of Web feeds (RSS, Atom). They have the advantage of being less influenced by spam and almost all dynamic websites provide such a feed, which is certainly not true for email alerts. Nevertheless, the use of Web feeds is hardly accepted by the general public; this may be due to the fact that you often need a separate feed reader. A Research Information Network study from the United Kingdom reports that the non use of RSS is even true for academic researchers [19].

Just like the use of social networks for finding information, the use of alerting services results in query formulation's becoming less important. However, alerting services also require extra skills and knowledge:

- ICT skills for installing and management of the right software or online tools (specifically for web feeds), and for subscribing and unsubscribing
- knowledge of relevant information sources that provide email alerts or web feeds and the skills to estimate the relevance, reliability and completeness of these sources
- the skill to recognise the information need while scanning the subjects or the headings of new messages. The reason why this can be problematic is discussed in the next sub section of this paper.

Another facet of personalisation in information research is the fact that search engines produce more 'customised result lists', based on the IP address of the computer that is used and the recent search history on that computer [20]. For mobile devices with GPS the location coordinates of the machine are of course used by the search engines to present location based personalisation of the search results [21]. A rather new phenomenon is personalisation, with content from your social networks ("social search", see for instance [22]). These examples make clear that it is really useful to have more than average knowledge of how search engines work and to have some skills to correct your search engine if that customisation is not desired.

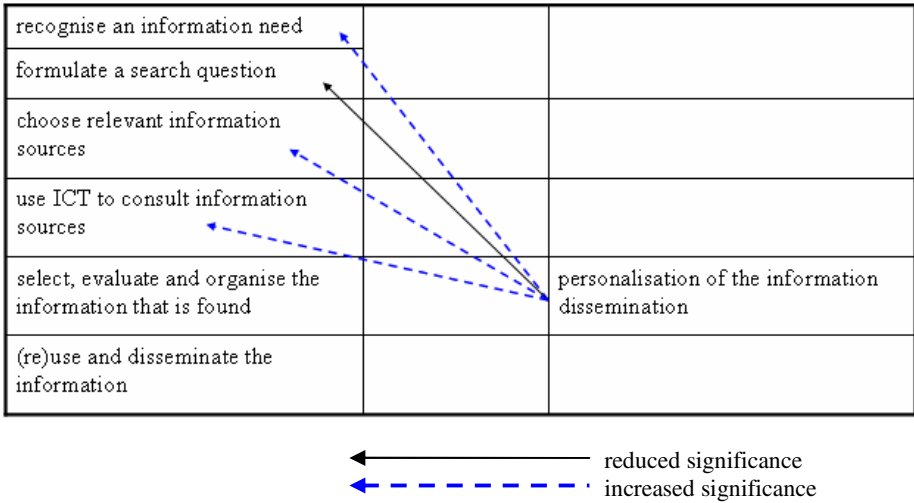


Fig. 4. Impact of the personalisation of the information dissemination on the Information Literacy sub skills

3.5 Fragmentation of Information

Discussing the use of alerting tools (web feeds, email) in the previous sub section of this paper, I observed that it is rather hard to recognise information needs while scanning the subjects or the headings of new messages. If someone follows fifteen weblogs, news sites or journals in an RSS aggregator like Netvibes or Google Reader, then it is quite normal that he or she receives about sixty new articles every day. To handle this daily amount of new information users only scan the headlines of the new items and restrict complete readings to the most meaningful messages. However, decisions about what is meaningful necessitate a lot of experience in evaluating the short headlines that are presented without their contextual information [23]. This fragmentation of information happens not just with alerting services but can also be seen in the presentation of the hit lists of the general search engines, since they only present titles of websites combined with the URL and a very short snippet of the phrase in which the keywords were found. In the section about easy to use publishing tools, it was emphasised that typical web 2.0 channels such as Wiki's and Weblogs demand extra skills from internet users to evaluate reliability, completeness and accuracy. The tendency to disseminate information on the internet separate from its context, underlines the growing importance of this Information Literacy sub skill.

However, the influence of information fragmentation goes even further. It also has the consequence that students and professionals have to piece together the information from a variety of information sources to their own meaningful knowledge framework [24]. In other words, the fragmentation of information also leads to an increasing importance of the sub skill to reuse and disseminate the information. David

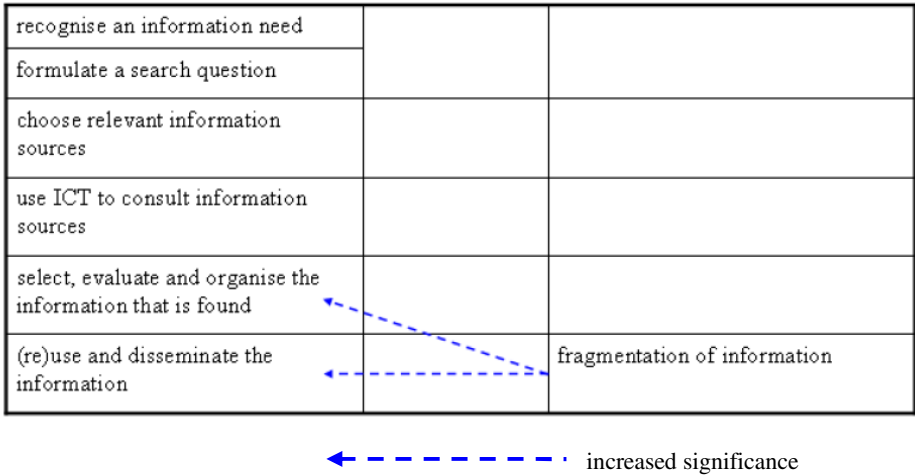


Fig. 5. Impact of the fragmentation of information on the Information Literacy sub skills

Weinberger refers to the same phenomenon when he claims that “Wikipedia [...] expects the reader to be *actively* involved, alert to the signs” [25].

4 Discussion and Conclusions

Figures 6 and 7 present a visual overview of the findings of this research. They make clear that some developments in information and communication technology reduce the significance of some more instrumental Information Literacy sub skills (Figure 6). This is true, for instance, for the formulation of search questions (“queries”) and for the use of ICT.

However, Figure 7 shows that the significance of some other Information Literacy sub skills has increased, not in spite of the use of advanced ICT tools but because of it. This is particularly true for the sub skills to evaluate and to reuse information. The fact that these higher order cognitive skills become more important due to the use of advanced technologies is the main finding after analysis of the literature that was found on recent developments in information and communication technology.

In this paper the focus was on the influence of ICT developments on the relevance of the concept Information Literacy. One of the findings is that the formulation of search queries has become less important because of the alternative methods for information seeking that are provided by various retrieval systems and the World Wide Web. This does not mean of course that the formulation of search queries is not important at all. What I wanted to emphasise is that the concept of Information Literacy is constituted by a set of sub skills and that extra attention should be paid to some more cognitive skills that refer to the use of information.

recognise an information need		easy access to bibliographic tools and full text documents
formulate a search question		
choose relevant information sources		easy to use publishing tools
use ICT to consult information sources		network functions of the WWW
select, evaluate and organise the information that is found		personalisation of the information dissemination
(re)use and disseminate the information		fragmentation of information

Fig. 6. Reduced significance of Information Literacy sub skills

recognise an information need		easy access to bibliographic tools and full text documents
formulate a search question		
choose relevant information sources		easy to use publishing tools
use ICT to consult information sources		network functions of the WWW
select, evaluate and organise the information that is found		personalisation of the information dissemination
(re)use and disseminate the information		fragmentation of information

Fig. 7. Increased significance of Information Literacy sub skills

This is in line with a change in higher education to “social constructivist learning” that we also know under names such as problem based learning, resource based learning and discovery learning. All these learning theories underline the importance of the exploration of resources and learning materials by the students themselves to build their own knowledge base [26]. Information problem solving plays an essential role in this educational approach and the necessary skills take a lot of effort and experience to master [27]. Answering the academic teachers who were paraphrased in the first section of this paper, it can be noticed that they focus too much on the instrumental sub skills of the competence Information Literacy and that they underestimate the complexity of the cognitive processes that also belong to it. However, this also means that Information Literacy programmes and courses should focus less on retrieval details and knowledge of library resources and should focus more on training students in

their writing skills and on the use of quality information in their school assignments. I guess that this can only become a success if library staff and teaching staff collaborate to incorporate these facets of the new knowledge work in the discipline based curricula.

References

1. Godwin, P.: Making Life Easier for Academics: How Librarians can Help Staff Weather the Technological Storm. *Journal of eLiteracy* 2, 68–79 (2005)
2. Boekhorst, A.: Informatievaardig Worden in Het Onderwijs, Een Informatiewetenschappelijk Perspectief: Een Vergelijkende Gevallestudie in Nederland en Zuid-Afrika. University of Pretoria, Pretoria (2000)
3. Webber, S., Johnston, B.: Conceptions of Information Literacy: New Perspectives and Implications. *Journal of Information Science* 26, 381–397 (2000)
4. Marton, F.: Phenomenography—Describing Conceptions of the World Around us. *Instructional Science* 10, 177–200 (1981)
5. Bruce, C.: *The Seven Faces of Information Literacy*. Auslib Press, Adelaide (1997)
6. Bruce, C.: The Phenomenon of Information Literacy. *Higher Education Research & Development* 17, 25–43 (1998)
7. Berkowitz, R., Eisenberg, M.: *Information Problem-Solving: The Big Six Skills Approach to Library & Information Skills Instruction*. Ablex Pub. Corp., Norwood (1990)
8. Maybee, C.: Undergraduate Perceptions of Information Use: The Basis for Creating User-Centered Student Information Literacy Instruction. *Journal of Academic Librarianship* 32, 79–85 (2006)
9. Bergman, M.: The Deep Web: Surfacing Hidden Value. *Journal of Electronic Publishing* 7 (2001)
10. Devine, J., Egger-Sider, F.: *Going Beyond Google: The Invisible Web in Learning and Teaching*. Neal-Schuman Publishers, New York (2009)
11. Godwin, P.: Information Literacy in the Age of Amateurs: How Google and Web 2.0 Affect Librarians' Support of Information Literacy, <http://www.ics.heacademy.ac.uk/italics/vol15iss4/godwin.pdf>
12. Morville, P.: *Ambient Findability*. O'Reilly, Sebastopol (2005)
13. Keen, A.: *The Cult of the Amateur: How Today's Internet is Killing our Culture*. Doubleday, New York (2007)
14. Johnson, L., Levine, A., Smith, R.: *The 2009 Horizon Report*. The New Media Consortium, Austin (2009)
15. Windham, C.: Getting Past Google: Perspectives on Information Literacy from the Millennial Mind. [S.L.], Educause, <http://connect.educause.edu/Library/ELI/GettingPastGooglePerspect/39341?time=1208003180>
16. Bates, M.: The Design of Browsing and Berrypicking Techniques for the Online Search Interface. *Online Review* 13, 407–424 (1989)
17. Carroll, J.: Expert Internet Information Access. *Journal of Educational Computing Research* 20, 209–222 (1999)
18. Mijnsbergen, E.: *Het Nieuwe Zoeken Heet Vragen*, <http://www.zbdigitaal.nl/2009/10/het-nieuwe-zoeken-heet-vragen.html>

19. Research Information Network: Researchers and Discovery Services: Behaviour, Perceptions, Needs,
<http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/researchers-and-discovery-services-behaviour-perc>
20. Garb, R.: More Transparency in Customized Search Results,
<http://googleblog.blogspot.com/2008/07/more-transparency-in-customized-search.html>
21. Arias, M., Cantera, J., Vegas, J.: Context-based Personalization for Mobile Web Search. In: VLDB, Auckland (New Zealand),
<http://persdb08.stanford.edu/5-PersDB-Paper.pdf>
22. Carmel, D., et al.: Personalized Social Search Based on the User's Social Network. In: Proceeding of the 18th ACM Conference on Information and Knowledge Management, Hong Kong, China, November 2-6 (2009)
23. Tuominen, K.: Information Literacy 2.0. *Signum* 5, 6–12 (2007)
24. Siemens, G., Tittenberger, P.: Handbook of Emerging Technologies for Learning. University of Manitoba, Manitoba,
<http://techcommittee.wikis.msad52.org/file/view/HETL.pdf>
25. Weinberger, D.: Everything is Miscellaneous. Henry Holt and Company, New York (2007)
26. Dochy, F., McDowell, L.: Assessment as a Tool for Learning. *Studies in Educational Evaluation* 23, 279–298 (1997)
27. Brand-Gruwel, S., Wopereis, I., Vermetten, Y.: Information Problem Solving by Experts and Novices: Analysis of a Complex Cognitive Skill. *Computers in Human Behavior* 21, 487–508 (2005)

Exploring the Competency of Academic Library Staff Engaging in Emerging Services*

Haruki Nagata**

University of Tsukuba, 5-3-614 Takezono 1chome, Tsukuba 305-0032 Japan
harungt@slis.tsukuba.ac.jp

Abstract. The basic capacity of library and information professionals is reexamined from two viewpoints. First, capacity is reshaped through emerging knowledge and skills under new information environments. Second, it is examined with more attention to the capability of high attainment of outcomes, i.e. 'competency', the core personality factors which predict job performance. This study focuses on 'competency' rather than knowledge and skill. In order to explore what competencies library and information professionals possess and utilize, the Behavioral Event Interview (BEI) has been administered to library staff (assistant librarians through section chiefs of libraries). Its transcript analysis confirms that interpersonal and motivational competencies are easily observable and rather dominant among three categories of competency. However the other category, (cognitive/intellectual competencies) was also identified through complimentary application of the Higher Education Role Analysis (HERA) instrument. And finally the study looks at how staff assess the required 'knowledge and experience' in their jobs and addresses the question of how knowledge (knowledge and skill) and core-personality (motive, traits and self-concept) competencies interact.

Keywords: Competency for library and information professionals, academic library, behavioral event interview, higher education role analysis.

1 Introduction

The knowledge and skills required of library and information professionals are changing. Body of Professional Knowledge published by CILIP (Chartered Institute of

* This paper is based on a research project sponsored by the Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science. The author would like to express gratitude to the sponsorship and colleagues of the project called LIPER2 (Library and Information Professions and Educations Renewal, 2006.4-2010.3).

** The author acknowledges for the courteous permission of HERA by ECC: Educational Competencies Consortium, Ltd. HERA scheme is an analysing tool for the activity of role holders as mentioned in 'HERA as a job analysis methodology', but the arguments here are not along the line of it. It should be noticed that this paper raises a different interpretation of HERA in a specific context. Further details of the use of HERA and its associated competency framework can be obtained from ECC's web site or by email via <mailto:contactus@ecc.ac.uk>

Library and Information Professionals) in 2004 [1], is a typical reframed system of competencies showing ‘core schema’, ‘applications environment’ and ‘generic and transferable skills’. But its perspective is confined to knowledge base. Different approaches [2], [3] appeared as early as 1997 and seem to offer a more extensive view. They refer to ‘competency’, identified by [4] as the underlying characteristics for “criterion-referenced effective and/or superior performance” in a job. Typically, however, these kinds of efforts have been derived not from an empirical approach, but from conceptual considerations in professional forums over the years [5]. Most were published in order to explore a new professional identity or various practical criteria in organizational assessment and staff training [6]. In a revised edition of *Competencies for Information Professionals of the 21st Century*, SLA (Special Library Association) [5], claimed to base these competencies on ‘best practice’ evidence though, its investigation has not been disclosed and evidence is not adequately confirmed. In addition the competency concept itself is rather ambiguous because it covers a wide range from inner traits of people to comprehensive professional knowledge and skills. For these reasons, in spite of their effort to be representative, the documents are an abstract expression and understanding them in working contexts is quite arduous.

This study aims to examine what competencies library and information professionals need, and how these competencies contribute to conducting their job activities. It focuses on the “personal competencies” [representing a set of attitudes, skills and values that enable practitioners to work effectively] rather than “professional competencies” [relating to the practitioner’s knowledge of information source, access, technology and management] described in the SLA document [5]. This approach has not obtained mass appeal in the library world up to now, as discovered in the review of the literature on competency and definition of the concept.

On the premise of taking an empirical approach to understanding the characteristic of ‘competency’, this study employs McClelland’s Behavioral Event Interview (BEI). The author visited five universities from 2007 to 2008 to interview library staffs. Transcript analyses of the interviews followed to pick up staff’s opinions of their jobs. Analyses were based on the methodology developed by Lyle M. Spencer, Jr. and Signe M. Spencer’s [7], [8]. Moreover this study adopts quantitative surveys and statistical verification as well.

This is a study to explore required competencies in human resource management of library and information professionals.

2 Exploring Competency

Competency is an individual’s underlying characteristics indicating their way of behaving and thinking. There are five types of competency characteristics: motives, traits, self-concept, knowledge and skill. These are also divided into two kinds; ‘*threshold competencies*’ (usually knowledge and skills) and ‘*differentiating competencies*’ (factors which distinguish superior from average performers) [7]. As a matter of first priority the threshold competencies, knowledge and skills, have received most attention so far. In recent decades people have been occupied with examining new types of threshold competencies because information and communication technologies have radically changed the workflows in libraries. As a consequence, the required competencies for library and information professional jobs had to be updated.

Research has considered competencies for library and information professionals. Nagata et al. [9] conducted 23 focus group interviews and a national survey of Japanese academic librarians from 2003 to 2005. Fifty-two items of knowledge and skills including emerging information and communication technologies were examined. Knowledge and skills pertaining to the services and communication were paid more attention than in a previous survey [10]. The research suggested that we should focus on staff attitudes and look at competency as a whole.

This study questions the *differentiating competencies*. Although knowledge and skills are quite easy to examine, the differentiating competencies are not. In order to factor out the core motives and traits this study has employed the Behavioral Event Interview (BEI) as a methodology. BEI is a structured interview that identifies people's recognition, actions and achievements in the job process in order to focus on the performance effectiveness of individuals. It gives "data about the interviewees' personality and cognitive style (e.g., what they think about, feel, and want to accomplish in dealing with the situation)" [7].

R.E. Boyatzis [11], who superseded McClelland's work, reviewed the competencies of superior managers across organizations using the BEI transcripts and presented a set of competencies of managers in 1981. Along with this attempt Spencer & Spencer looked at competencies found in reports of more than 200 diverse jobs. Those reports contain clusters of distinguished competencies, namely characteristics of superior performers identified by BEI studies. And such competencies are shown in narrative definition with "specified behavioral ways of demonstrating the competency in the job". Spencer & Spencer examined them based on behavioral indicators ("lowest common denominator, or smallest unit of observation directly comparable across all models") and made up a list of approximately 769 behavioral indicators. Then finally 360 behavior indicators were extracted as items of a preliminary dictionary [7].

"The competency dictionary presents competencies in generic form, in scales designed to cover behavior in a wide range of jobs" [7] which is comprised of 19 cluster concepts (Achievement Orientation; Concern for Order, Quality, and Accuracy; Initiative; Information Seeking; Interpersonal Understanding; Customer Service Orientation; Impact and Influence; Organizational Awareness; Relationship Building; Developing Others; Directiveness [Assertiveness and Use of Positional Power]; Teamwork and Cooperation; Team Leadership; Analytical Thinking; Conceptual Thinking; Technical/Professional/Managerial Expertise; Self-Control; Self-Confidence; Flexibility). The cluster concept represents the underlying intent to cope with the situation [7] and provides a basic unit for setting an individual job competency model. Table 1 shows the

Table 1. Spencer & Spencer's Generic Competency Model for Technical Professionals

Weight	Competency clusters
6	Achievement Orientation
5	Impact and Influence
4	Conceptual Thinking; Analytical thinking; Initiatives
3	Self-Confidence; Interpersonal Understanding
2	Concern for Order; Information-Seeking; Teamwork and Cooperation; Expertise
1	Customer Service Orientation

competency model for technical professionals whose work involves “the use of technical knowledge”. As far as library and information professionals are concerned with tasks of technical processing they can be seen to belong to this group. The weight in the left column refers to the “relative frequency with which each competency distinguishes superior from average performers” [7]. According to this table, technical professionals seem to be more concerned with their achievement and capabilities to realize their aims than with attention to customers.

3 Behavioral Event Interview (BEI)

3.1 Samples for BEI

From January 2007 to February 2008 the author interviewed 27 library staff at five very vigorously active university libraries in Japan. As summarized in Table 2, twelve of the participating staff are assessed as superior performers by their managers based on their contribution to the organization, seven are average performers and eight are managers. Managers were interviewed whenever possible, since their opinions are adopted to judge the staff.

The superior performers are working at middle level positions and their ages range from late 20s to 40s. Gender ratio is almost the same, with a few more females than males. These performers are mostly engaged in emerging services, e.g. institutional repository management, user interface production of library information services, customer service for digitized materials, organizing digital materials, and information literacy education.

Each BEI took about an hour to conduct, and was recorded by a digital IC (integrated circuit) recorder. The question topics mostly covered the job success and failure stories with the introduction of job details. The transcripts of the interviews were then prepared for analysis.

Table 2. Categories of library staff interviewed by the BEI

	BEI		Managers
	Superior performers	Average performers	
A Library	2	2	3
B Library	2	2	
C Library	2		1
D Library	4	1	3
E Library	2	2	1
Total	12	7	8

3.2 Findings from BEI

The transcripts of interviews were analyzed within the scope of Spencer & Spencer’s specification. To do thematic analysis to identify themes and patterns, phrases relating to ‘Situation’, ‘Who is involved’, ‘Thoughts’, ‘Feelings’, ‘Actions’, ‘Outcomes’, ‘Other characteristics (Physical appearance and so on)’ were extracted [7]. Figure 1

shows some examples illustrating the themes of ‘Situation,’ ‘Who is involved’ and ‘Motivation’. These are statements differentiating superior from average performers.

1 状況(Situation)

1-1 利用者ってやっぱり、知らないっていうか、せっかく、こんなにいいものがあるのに、使いこなせてないですし、すごく高いお金で買っているいろんなもの図書館にあるんですよね。それを、なんか、やっぱりつないで、つなぐ仕事っていうのが、それをつなぐ仕事っていうのは私たちしかないですよね。
(状況と役割の認識)

Translation: 1-1 Users do not know library collections so much; actually they cannot utilize such fantastic goods. Our library has paid lots of money to collect wide variety of materials. It is only we that link users to those collections. (Recognition of role)

1-2 やっぱり、分業っていう意識が大きいので。なるべく自分では、もちろんやっている仕事は、一部分なんですけど、なるべく全体を見たいなと思いつつながら、ええ、ひそかに、ああ、ここはこうだいいなと思いつつながら、なんかのときには、発言したり行動したり、っていうことをしたほうがいいのかなと思いつつ。わりあいすぐ上の上司には、たぶん私はいうほうだと思います、いってどうなるという、あれじゃないんですけど、とりあえず、いってこうかな、という気持ちがあります。」(組織状況の把握)

Translation: 1-2 We run our business on the basis of division of labor. My work is a part of it, however I would like to look at it holistically. And it might be better for me to mention issues and get in action when necessary. My superior takes me for a valuable person. I think I should speak out, but don't expect the result of it. (Understanding of organization context)

2 関係者(Who is involved)

2-1 その、なんか、自分がやっていることが、周りの人に知ってもらえたり反応があったりで、それで、その人たちとコミュニケーションを取っているうちに、自分自身の考えがこう、まあやることがその、少しずつ、洗練、とかいう話、そこまではないかもですけど、幅広くなる、っていう部分があるのかな、と。(同僚の存在)

Translation: 2-1 Let me think, once I have a reaction toward my work from the people around. And I have kept in touch with them about my work, and then owing to communication with them I was able to hit on the idea of sophisticating, enlarging, widening my concept, in other words what I am tackling. It is not clear, though. (Development with colleagues)

2-2 あの、…なんていうのかな、自分はただ純粋にやっているんだっていうところをみせることで、まあ、だんだん、ひとつふたつやっていると、みんながあの人だったらつくってくれるんじゃない、みたいな感じに、雰囲気になってきた。(同僚からの目)

Translation: 2-2 Well, how can I say? That I was doing my sheer volume of work and obtaining results was appealing until my colleagues change their attitude and started singing another tune about me. They now think that he(I) will deliver it for them. (Collegiality)

2-3 やっぱりその、そこで何人かのコミュニティをつくって、ぐーっと引張って行って、そこに、その、もしそれがよいものであれば、そこにこう、周りに、ユーザとかですかね、そこに賛同する人が集まる、みたいな。結局そういうことなのかな、と。自分で一人でやっていることが負担じゃないっていうのは、あ、どこもそうなのね、っていうところがみえてきた。(仲間の広がり)

Translation: 2-3 All in all, I try to make a circle of colleagues, and lead them aggressively to an objective. If it is OK with others, it has been possible to collect the advocates around me. In effect I realized I am not the only one who took it on. (Peer identity)

4. モティベーション(Motivation)

4-1 だから絶対成功する、っていう思いでやっている。(中略)失敗したら、後がないわけじゃないんだけど、失敗するようなものは絶対つからない、という思いで。だからツメを大事にしている。まあ、人からいわせれば細かすぎる、というはあるんですけど。(達成の重視)

Translation: 4-1 I have tackled it with strong confidence to make it. *snip* Even though it failed I still have chances to recover, but I am filled with a conviction that I won't fail. I think a great deal about ending of each task. Maybe I worry too much about details. (Achievement-centered)

4-2 たぶん ひとつに理由を集約するなら、その、プロ意識。自分の思い描いている図書館員ってどんなだろう、やっぱり、外国人がきてもちろんと対応したいよね、っていう。だから、[語学研修] 自分の場合海外旅行に行くためではなかった。(仕事の達成)

Translation: 4-2 Perhaps, speaking with one voice, it is a kind of professionalism. What kind of librarian should I be? When encountering the foreign customer, I would like to render a proper service. So language program is a must for my job, not for oversea trip on holidays. (Job Attainment)

Fig. 1. Part of BEI transcript analysis

Each additional phrase within parentheses in the end is the author's annotation. Checking the competencies of interviewees, the phrases were then categorized into three general categories; 1) Cognitive/Intellectual competencies ("skills involved in creating, getting, or using information; learning from experience; objectively analyzing data; or thinking through alternatives for action, logical thought, and divergent thinking"), 2) Interpersonal competencies ("skills involved in communicating with, understanding and influencing others, such as accurate empathy, positive regards and expectations, articulateness and public speaking ability"), 3) Motivation competencies ("the needs or drives that cause people to want and do different things; achievement, affiliation, and power motives; and emotional self-control") [7]. From these examined transcripts vocabularies for competency of the library and information staff emerged.

The examples in Figure 1 can be classified exclusively into three categories, 1) Cognitive/Intellectual competencies [illustrated by 1-1], 2) Interpersonal competencies [by 1-2, 2-1, 2-2 and 2-3], and 3) Motivation competencies [by 4-1 and 4-2]. Findings reveal that superior performers certainly talk about their competencies to communicate with colleagues and the occurrence of interpersonal relationships is rather high among them. It is quite understandable for them to focus on the interpersonal relationships with intent to perform their tasks with felicity in their organization and know how to express themselves to others. Of course they have highly motivated attitudes which are also frequently found in the transcripts. Evidence among superior performers for cognitive/intellectual competencies however is less apparent in their direct statements.

4 Complementary Survey

4.1 HERA as a Job Analysis Methodology

HERA (Higher Education Role Analysis) was designed and developed as the job evaluation scheme by Educational Competencies Consortium Ltd. [12]. Universities in the UK introduced it as one of the measures of job analysis for the reform of pay structures. Many institutions now use HERA to assess jobs and rank all kinds of posts into one salary scale. In using the HERA, trained role analysts interview persons who gave a reply to the questionnaire about their job. They evaluate jobs by checking the scored evidence of role requirements in responses to the questionnaire. So first staff have to answer the questionnaire which consists of 14 elements from 'Communication' to 'Knowledge and Experience' shown in Table 3. Each element has a series of 50 questions which draw out evidence for what roles are required of the positions. This study's concern is limited to staff answers to certain questions, and does not extend to the scoring and ranking procedure by analysts.

The HERA questionnaire sets a generic perspective to analyze jobs rather than looking into specific requirements of professional knowledge and skills. It covers all areas of competency employed in various positions in higher education institutions. So it can be hypothesized that staff answers could show significant differences in job interpretation between the superior and average performers to a certain extent if comparing the answers of those who are engaging in same or similar tasks. In fact HERA includes two types of questions aimed to show the divergence due to the respondents'

Table 3. Categories of library staff interviewed by the BEI

1) Communication (covers communication through written, electronic or visual means and oral communication, formally and informally)
2) Teamwork and Motivation (cover team work and team leadership when working in both internal and external teams)
3) Liaison and Networking (cover liaising with others both within and outside the university and creating networks of useful contacts)
4) Service Delivery (covers the provision of help and assistance to a high standard of service to students, visitors, members of staff and other users of the university)
5) Decision Making Processes and Outcomes (cover the impact of decisions within the institutions and externally)
6) Planning and Organising Resources (cover organising, prioritising and planning time and resources, be they human, physical or financial)
7) Initiative and Problem Solving (cover identifying or developing options and selecting solutions to problems which occur in the role)
8) Analysis and Research (cover investigating issues, analyzing information and carrying out research)
9) Sensory and Physical Demands (cover the sensory and physical aspects of the role required to complete tasks)
10) Work Environment (covers the impact the working environment has on the individual and their ability to respond to and control that environment safely)
11) Pastoral Care and Welfare (cover the welfare and well being of students and staff within the institution in both formal and informal situations)
12) Team Development (covers the development of the skills and knowledge of others in the work team)
13) Teaching and Learning Support (cover the development of the skills and knowledge of students and others who are not part of the work team)
14) Knowledge and Experience (cover the relevant knowledge needed to carry out the role, however acquired, whether this is technical, professional or specialist)

opinion or way of thinking. One type includes the questions covering a wide scope of issues, namely '14) Knowledge Experience' and '13) Teaching and Learning Support.' The other type includes the questions which need some consideration and judgment when answering them, such as those under groupings of '5) Decision Making Processes and Outcomes', and '7) Initiative and Problem Solving'. This study has employed this part of the HERA as a complementary instrument to check staff interpretation of their job role in their library. Incidentally HERA has also enough elements to illustrate the role of interpersonal activities including the channel (written, electronic and visual, oral) and ways (liaison, networking and teambuilding) of communication, which, however, are not used here. The sample population for conducting the HERA survey is the same as used for the interview shown in Table 2 (excluding managers).

4.2 Some Features of Superior Performers Extracted by the HERA

Several characteristics of superior performers were identified from responses to the HERA questionnaire. One example is found in responses to the HERA question 30, illustrated in Figure 2. This question asks whether the respondent is "required to resolve problems where there is a lack of precedent which calls for innovation and creative thought to develop appropriate options; anticipate problems and make projections; initiate solutions which do not limit future choices," namely how to handle contingent situations. Naturally the staff needs cognitive/intellectual competencies as well as good motivation competencies to cope with such situations. This type of question is a way to uncover how well the cognitive/intellectual competency is possessed by role holders.

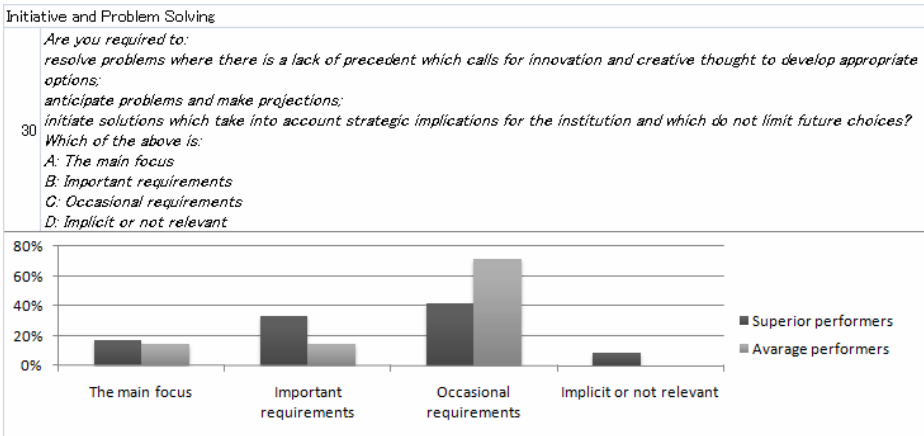


Fig. 2. Comparison 1. Requirement level of Problem Solving.

As a result fifty percent of superior performers chose ‘The main focus’ or ‘Important requirement’ toward the question. On the contrary, seventy-one percent of average performers picked ‘occasional requirements’. Certain differences between superior and average performers’ responses are quite visible. Half of the superior performers seem to have a more active self-perception to tackle contingencies; on the other hand a majority of average performers may be reluctant to identify themselves as doing so. Superior performers can be recognized by their response to take a positive position with some competencies. However, converting each option of A through D into 1, 2, 3, 4 respectively, the result of an independent sample t-test of mean values does not show a statistically significant difference in responses between superior performers and average performers ($p=.71 > 0.05$, the difference of mean: .155, Variance: .708). Maybe the implication of this question is too broad or the sample size is too small to identify a clear statistical difference.

Figure 3 shows the responses to HERA question 47 inquiring whether the respondent is “required to teach or train students or others on specific tasks, issues or activities; assess performance and provide feedback during the event.” This question relates to how to develop the person who is not a colleague. This task may be recognized as ‘user education.’ It is a comparatively new area among library and information professional duties, so positively answering some questions poses a kind of challenge (A: Does the role holder develop innovative approaches to the learning experience and the curriculum; originate content and methodology? and B: Does the role holder design content or learning materials within existing frameworks; make appropriate modifications to existing materials on the basis of the knowledge or experience of the learner(s)?). Average performers responses are formulaic or ‘not relevant’.

Question 47 shares the same issues with question 30. Obviously the characteristics which these questions uncover are oriented to the cognitive/intellectual competencies as well as motivation competencies. But the role of this question is to clearly articulate the different attitude of average and superior performers. The gap between the two types of performers is easily sizeable as shown in Figure 3. And statistical analysis (independent sample t-test of mean values) also shows a significant difference between superior performers and average performers ($p=0.017 < 0.05$).

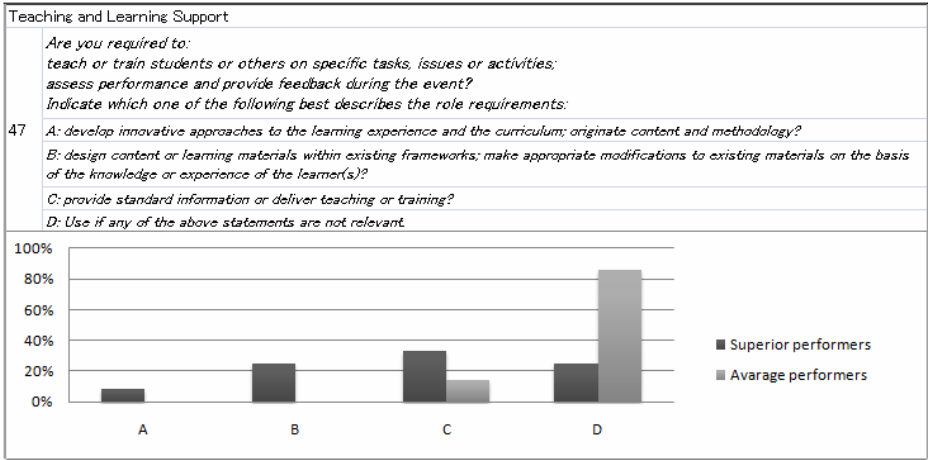


Fig. 3. Comparison 2. Solutions based on Knowledge and Skills.

Finally it might be worthwhile to look at how necessary library staffs think their knowledge and skills are to performing their jobs. Figure 4 illustrates the responses to question 50 (Knowledge and experience). About 27-28% of all respondents think that their required knowledge should be at essential levels but the rest of them think they would like to utilize higher knowledge and experience in their work and intend to refresh themselves as professionals. This result interestingly does not show a clear difference between superior and average performers. The finding might be interpreted that knowledge and skill are recognized as necessary competencies, referenced as ‘threshold competencies’, which is important to understand job competencies.

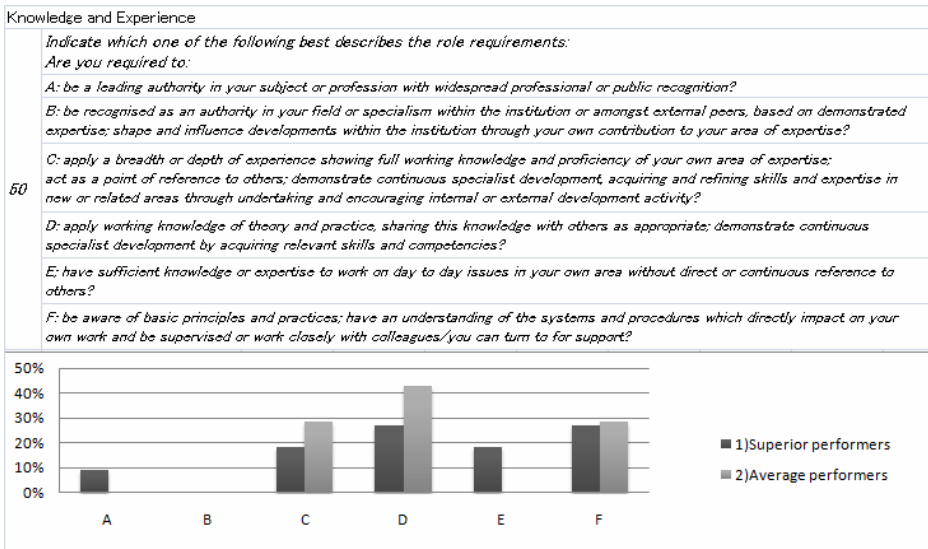


Fig. 4. Comparison 3. Evaluation on Knowledge and Skills.

5 Summary and Discussion

Anne Bell, the university librarian at University of Warwick, UK asserted that we were focusing too much on the process of professional work (personal communication, February 5, 2008). She raised a problem of library professionals who are sticking to their acquired competencies for their routine processes and hesitating to change their tasks. Now reviewing library and information professionals' transforming competency from new perspectives is a vital issue. For the purpose intended, this study has tried to reexamine the competency of library and information professionals from the perspective of creative attitudes rather than the attention to the process of their jobs.

BEI and HERA were used to identify library and information professionals' competencies. The superior performers have shown their excellent interpersonal skills and high motivation during their interviews. One might conclude that the competencies of library and information professionals have a more customer-oriented profile. Their characteristics are different from the competencies of 'Technical Professionals', whose priorities are more oriented to achievement/impact and cognitive abilities in their professional field as shown in Table 1. In addition some cognitive/intellectual competencies of the superior performers were easily confirmed by HERA survey results. This trial's scope is not broad enough to set up a competency dictionary/model for library and information professionals in general. But certainly the study uncovered some characteristics to develop a typography of competencies of academic library staff engaging in emerging services.

As a matter of course, threshold competencies (knowledge and skills) are essential to performance. Lately these have been modified a lot and continue to change swiftly, responding to social and technological development. For that reason, researchers should devote attention to understanding them. Challenges remain to associate knowledge and skills work with best practice, and especially methods to produce efficient processes. Cultivation of core-personality competencies is one answer McClelland has suggested. But future research might identify more concrete ideas to improve preparation for the changing nature of librarianship and other information professions. So the author advocates the association between knowledge (knowledge and skills) and core-personality (motive, traits and self-image) of "competency" and seeking to identify ways for these to work together to improve staff performance.

At the moment, the author places his hope on the formation of 'communities of practice' which is a concept developed by Jean Lave and Etienne Wenger [13]. It is the process of sharing knowledge and skills within a group or community. Actually the members of the community learn from each other and have an opportunity to develop themselves not only through knowledge but also in core-personality competencies. In this study as well it was visible that most superior performers have this kind of communication (shown in Figure 1) in their work places including networked communities.

Acknowledgments. Acknowledgement is made especially to the twenty-seven persons who were willing to be interviewed and responded to the survey. And I would like to thank Danuta A. Nitecki, Dean of Libraries at Drexel University who gave me concrete suggestions.

References

1. Chartered Institute of Library and Information Professionals, <http://www.cilip.org.uk/sitecollectiondocuments/PDFs/qualificationschartership/BPK.pdf>
2. Marshall, J.: Competencies for Special Librarians 1997: Competencies for Special Librarians of the 21st Century. SLA Board by the Special Committee on Competencies for Special Librarians (1996)
3. McNeil, B., Giesecke, J.: Core Competencies for Libraries and Library Staff. In: American Library Association. Staff Development: A Practical Guide, 3rd ed., pp. 49–62 (2001)
4. McClelland, D.C.: Testing for Competence Rather than for “Intelligence”. *American Psychologist* 28, 1–14 (1973)
5. Abels, E., Jones, R., Latham, J., Magnoni, D., Marshall, J.G.: Competencies for Information Professionals of the 21st Century. Revised edition (2003)
6. Soutter, J.L.: Librarian Competency: a Description of Trends in the Peer-Reviewed Journal Literature of 2001–2005. *Partnership: The Canadian Journal of Library and Information Practice and Research* 2 (2007), <http://journal.lib.uoguelph.ca/index.php/perj/article/view/252>
7. Spencer Jr., L.M., Spencer, S.M.: *Competence at Work: Models for Superior Performance*. J. Wiley, New York (1993)
8. Getha-Taylor, H.: Collaborative Competencies; for Presentation at the Public Management Research Conference (2007)
9. Nagata, H., Toda, S., Itsumura, H., Koyama, K., Saito, Y., Suzuki, M., Takahashi, N.: Body of Professional Knowledge Required for Academic Librarians in Japan. In: Khoo, C. (ed.) *Proceeding of the Asian-Pacific Conference on Library & Information Education & Practice 2006 (A-LIEP 2006)*, pp. 316–327 (2006)
10. Miura, I., Kikuchi, S., Mori, T., Horikawa, T.: Daigaku Toshokanin No Chisiki Besu to Tshokan Gaku Kyoiku (Knowledge Base for Academic Librarians and Library Science Education). 1-2. *Annals of Japan Society for Library Science* 37, 49–63, 103–116 (1991)
11. Boyatzis, R.E.: *The Competent Manager: a Model for Effective Performance*. J. Wiley, New York (1982)
12. Higher Education Role Analysis (HERA) and Job Evaluation, <http://www.admin.cam.ac.uk/offices/hr/reward/hera.html>
13. Wenger, E., McDermott, R.A., Snyder, W.: *Cultivating Communities of Practice: a Guide to Management Knowledge*. Harvard Business School, Boston (2002)

Programming New Learning Spaces: The Changing Nature of Academic Library Buildings*

H. Lea Wells and Jordan M. Scepanski

Jordan Wells Associates, 908 Arrowhead Road, Chapel Hill,
North Carolina 27514-3902, U.S.A.
jordanwells@me.com

Abstract. Greater attention to student and faculty expectations and the impact of technology on library and information operations are combining to significantly alter library physical facilities and programming. Academic institutions undertaking new library construction or renovating older buildings are rethinking how library space is used and configured. Alliances and partnerships among a variety of academic and support units are resulting in facilities more responsive to users' work habits and their preferences for comfortable, inviting, and productive environments. This paper reports on a study of new library and learning spaces conducted for a major American university planning a campus in the Middle East.

Keywords: Library buildings, academic collaboration, academic partnerships, new libraries, renovated libraries, library 2.0.

1 Background and Methodology

In the winter of 2008, at the request of a major private American university undertaking establishment of a campus in the Middle East, a concept paper was developed to guide planning of a new library, and in particular its physical spaces. The investigation placed emphasis upon identification of programs and services that might be included in, or adjacent to, the library facility, whether or not these ultimately were to be incorporated into the library's administrative structure. In preparing the report the consultants conducted physical library and internet research; reviewed planning documents of the parent university and of peer institutions; interviewed individuals with experience in creating new library programs, planning new library structures, or renovating older facilities; and generally assessed issues of space, staffing, and culture. Extensive examination of the websites of universities established in the past decade and a half in the United States and in the Middle East was undertaken. In addition, library programs at a number of liberal arts institutions and research universities

* This paper is based upon an investigation conducted on behalf of New York University in 2008 as it undertook planning of a new campus in Abu Dhabi, capital of the United Arab Emirates. Financial and other support of NYU and its Bobst Library staff in the research and preparation of that 2008 report is gratefully acknowledged.

were considered, especially those that had undergone construction and renovation in recent years. Readings included literature on the changing nature of student study, new developments in faculty teaching and research, learning spaces, campus collaborative endeavors, and innovative information technology implementation in library settings. Visits were made to gather additional data from libraries that recently engaged in construction/renovation projects and information was obtained through telephone conversations or e-mail communications with individuals knowledgeable about new libraries, or about new approaches to library service. What follows is a summary of the study's findings.

2 Findings and Discussion

Academic institutions are making significant changes in their library programming, changes enabled by technological developments but driven by desire to better meet the learning, teaching, and research requirements of their clientele. These changes result in new physical facilities or renovations to create environments more conducive to the ways the current generation of students learn and live, and to how faculty best teach and do research. Programming is being implemented to complement the new spaces and structures, very often developed in a collaborative fashion with other campus organizations. The alliances forged most frequently have been with information technology partners, but more and more libraries are directly associating their programs with student and faculty support entities and food service operations in an attempt to reach out to more broadly-based constituencies and to break down what are viewed by some as artificial administrative barriers.

2.1 Space Planning for Libraries

Previously an estimation of overall space requirements for a new academic library would first take into account the number of book and serial volumes that would have to be accommodated in the facility being planned. With a standardized formula calculating square footage based on expected volume count, certain percentages of that space allocation would be used to determine user and service area sizing and then the anticipated numbers of different types of staff would drive final footage for back office or work areas.

In an era when significant access to information in electronic formats makes it possible to provide an effective library program without having to invest heavily in print resources, space planning for libraries becomes more complex. Using older academic planning tools with smaller physical collections results in less space for library users and smaller areas in which to serve them. As importantly, standard formulas have not caught up with the range of programs and services being incorporated into modern library facilities, nor do they take into account other campus units and operations that have begun to carry out many of their activities in the library building.

Students seek comfort and convenience in their library experience. They want the library to be a place to retreat to, a sanctuary where they can work free from the

distractions of home or dormitory, but also a setting in which they can engage in group study and collaborate on project assignments. They want the facility available at most hours of the day or night; they want a place in which they can eat, socialize, rest and relax, and yet one whose atmosphere is conducive to serious scholarship. When they seek learning or research assistance, they do not understand boundaries created by organizational structures and as much as possible want to be able to go to one place for most of their service needs.

Today's students reside in a virtual world as much as in the "real" one [1], and given the technologies they have access to, they see little limitation imposed by walls. They also take a collaborative approach to learning, well beyond anything exhibited by previous generations. This is the result of pedagogical shifts and demands in the workplace requiring teams and partnerships, but also because students like functioning in this fashion. They are at home in large rooms with lots of people, but they desire smaller settings and quieter places too. Students also are seeking opportunities to establish relationships with their instructors outside of the traditional classroom setting. The new library is a primary place where less formal interaction between students and faculty can occur.

With faculty, expectations of the library are more complex. Their views are being affected by educational and societal changes. For many faculty members the library exists only on computer screens when they are pursuing their research, and most are highly positive about that online presence. But many also see the library as a social and professional arena where teaching can take place and where important collaboration can go on. The centralized libraries of today have been compared to the discipline-based departmental or college libraries of old, where faculty could relax, connect with colleagues, pursue collaborations, and peruse the journal literature [2]. Faculty are finding in modern day library settings a social ambiance and interaction that is valuable and attractive. And this is taking place even as the primary purpose for visiting a library facility in the past—searching its scholarly information holdings—has been made less necessary by technology. The library is seen as a natural place for cross-disciplinary connections and conversations to occur.

Student fluency in a range of learned skills, with words, images, music, etc. in information gathering and in presentation techniques, is important. Libraries, through partnerships with faculty, information technologists, and student services professionals, and by offering equipment and settings to facilitate acquisition of these skills, contribute to well-rounded educations and to life-long learning abilities. Teaching students how to approach a research topic, how to find resources relevant to a need, how to assess reliability and currency of what is found, and how to appropriately attribute sources is now a hallmark of academic librarianship. Increasingly, being able to effectively convey one's work in written formats, or through sound and images, is demonstration of educational achievement.

2.2 The Changing Face of Academic Libraries

In the interest of better assistance for students, some libraries are consolidating desks where in the past different types of services were offered. Most often this takes the

form of integrating circulation and reference services, but it now is not uncommon to find a greater variety of first-level staff at the same place lending books and laptops, answering directional questions or referring substantive inquiries, and otherwise handling very effectively much of the normal traffic that most service points in a traditional library encounter. Some libraries have eliminated reference desks or at least have reassigned the librarians who staffed them. Electronic inquiries do not require a public desk and librarians increasingly carry out this work in classrooms, in academic departments, or online.

Most academic libraries today are investing heavily in digital reference tools and in robust collections of electronic journals. Some have decided to subscribe to only a bare minimum of print periodicals. The University of California at Merced library, with broad access to the resources of the University of California system and the California Digital Library, spends only about a thousand dollars a year on a hundred or so hard copy magazines, popular titles chosen by the students and minimally managed (few records are kept for them and when tattered or torn they are simply discarded) [2].

Round the clock operations during much of the school week at libraries of larger universities, and at many smaller institutions as well, is a student expectation. And as more campus services are provided in the library facility, there will be greater demand for extended hours.

A collection of materials typically found in libraries supporting elementary and secondary educational programs, or in the education departments themselves, are curriculum-based resources relating to primary- and secondary-level teaching. Encompassing textbooks, children's books, lesson plans, teaching tools, equipment for constructing classroom aids, and all manner of material useful for teacher training and professional development, these curriculum resource centers are essential adjuncts to strong programs in education. Given significant interest in elementary and secondary school reform, such centers could be most important in training teachers and in furthering professional development.

Among the services and facilities increasingly being found in libraries are those having a technological focus, particularly where assistance to students and faculty is involved, and those providing food and refreshment.

Student computing laboratories, once almost exclusively the domain of campus information technology units, and physically found in either consolidated computer services facilities or scattered about in academic departments, now are commonly situated in library buildings, usually through some sort of public services partnership between information technology personnel and library staff. This development began at least twenty-five years ago as the value of locating a high-use and staff-intensive operation in an existing extended-hours, service-oriented program became obvious. Today one would be hard pressed to find a respectable academic library without some type of computing laboratory on its premises.

Space given over to workstations that provide access to the full range of information resources either owned by a library or available through its gateway are now

found everywhere under the term “commons,” as in “Information Commons,” or “Learning Commons.” At times the word “collaboratory” is used. Where traditional computer labs largely have been seen as places for production of products—term papers written using word processing software and presentations for class assignments—or for manipulation and analysis of data through spreadsheet or statistical packages, “commons” and “collaboratories” are meant to suggest places one goes to gather information and to work together. In their best forms these service areas offer both technical and informational assistance in an integrated and user-friendly fashion. An example is at the Georgia Institute of Technology [3] Library.

Smaller study rooms have been fixtures in academic libraries for many decades. They always have been popular venues for groups and those carrying out team assignments. What has changed in recent years is the need for such rooms to be outfitted at a minimum with network and power outlets and preferably with display screens to which laptops can be connected. These rooms range in size from seating for two to four individuals to as many as a dozen or so. The North Carolina State University [4] Library offers such.

Another type of room in new or renovated library facilities, one requiring fairly sophisticated technological components, is intended for practice of presentations using multi-media resources. These presentations might be by groups or individuals, for a class exercise, or even for a job interview. They usually require a degree of technology greater than that found in the outfitted group studies described above. Some examples of these rooms are found in the libraries at Portland State University [5] and at Mississippi State University [6].

As previously noted, information literacy (also referred to as fluency or competency) is considered an essential attribute of an educated person and academic libraries have embraced a leading role in teaching students how to go about accessing and assessing scholarly resources. In doing so librarians require classrooms and training facilities equipped to demonstrate online information sources and to give students hands-on practice in finding material. Such facilities have been added to most libraries over the past decade.

As universities recognize the need to make available new teaching settings and experiment with “smart” classrooms, libraries are natural, relatively neutral places to locate them and to assure they get maximum use and receive necessary maintenance. While these classrooms can be installed anywhere on campus, placement in the library building would seem to be a cost effective approach to implementing these facilities. The University of Oregon [7] Libraries offer such classrooms.

Multi-media or digital media production labs have found their way into modern libraries as well. Much academic work today—student projects, development of faculty teaching tools, content compilation, analysis and annotation, and other digital scholarship—requires conversion of materials from one format to another, or a mixing of media in sometimes sophisticated ways. Computing, scanning, editing, copying, and printing equipment, for sound and images—both still and moving—and attendant software, are necessary components of such labs. The North Carolina State University [8] Libraries have such a lab.

The technological programming of many of today’s libraries described above is in addition to workstations that might be clustered about the various rooms and floors to

enable distributed information access. The extent to which workstations may need to be provided depends upon whether students are likely to have their own laptops (or be required to have them).

Just about every library renovation today includes some sort of coffee shop or refreshment area in a prominent place in the facility. Many attribute this phenomenon to the popularity of food and drink operations in large book and music chain stores. If these commercial entities are experiencing success in attracting new customers by selling lattes near the book stock, the reasoning goes, why not do something similar in libraries. Whatever the origin or reason, it is indisputable that the long-held tradition of no eating and drinking in the library facility is dead. Increasingly libraries at academic institutions offer refreshment stations of some kind. The Harvard University [9] Lamont Library Café is one of many such examples.

2.3 Academic Services Traditionally Not Located in a Library

The programs and services outlined in the preceding paragraphs, while by no means available in every academic library, are found in many, and especially in those that have had opportunity to reconfigure and renew their facilities. But other student services have begun to appear too, not as often as those that are technology-centric, nor as popular as coffee shops, but finding a presence nonetheless. The reason is an understanding of a mutual orientation toward student service and recognition that one-stop shopping, or at least provision of access to many campus activities in or near one building, can enhance the student experience.

Among student services types of programs located in or proximate to some libraries are academic advising, bookstore operations, bridging programs, career services, counseling, language learning, publishing, and writing centers.

Advising offices are sometimes co-located with the library or advising outposts are established there. The Wayne State University [10] academic advising center is located in its undergraduate library.

Bookstores are seen by some to be natural neighbors for libraries. The University of California, Merced [11] bookstore is in the library building.

At universities that do most of their teaching in English, but who have students whose primary language is something else, readiness or remedial language instruction is usually essential. A helpful service and a good match for location in the library building is a language resource center. For an institution with a language preparatory program, graded language readers and audiovisual language learning tools are appropriately provided through the library. Laboratories for native English speakers studying foreign languages also have found their way into some libraries. The Harvard University [12] Lamont Library offers this type of resource center.

With collections of background material on potential employers of a university's graduates, and books and articles on application and resume preparation and interviewing, career counseling centers have found comfortable homes in libraries. The availability of meeting rooms that can be used for interview sessions also is a positive aspect of such a location. The Florida State University [13] Library houses a career center and offers virtual connections to materials in-house and beyond.

Physical placement of psychological counseling services on campus can be a delicate matter. These programs aim to assure their clients anonymity and often attempt

to have office locations that shield student entrances and exits. While one cannot say that library locations are common for this type of service, there are examples of collaborative programming that bring counseling workshops into the library facility [3]. Georgia Tech is said to have experienced a four-fold increase in counseling center workshop participation after they were offered in the library. Students reportedly perceived the counseling facility as a “sick” place, whereas the library was seen as “normal” and “healthy” [14]. The University of North Carolina at Charlotte [15] counseling center has a long history of co-location with the library.

Administrative linkages between libraries and university presses are well established at some institutions. For example, Wayne State University has merged library and university press reporting.

Finally, helping students improve their writing skills can be considered a “library-like” activity. Writing centers are a natural fit with libraries, often linked to an information commons. Some institutions hold writing workshops or have writing staff in the library at certain hours, even when the actual center is housed elsewhere. For example, the Arizona State University [16] Hayden Library provides space for use by writing center staff a few hours each week. A good example of a library-based writing center is at [17].

2.4 Other Academic Services

Beyond direct service to students, there are three other programs whose relationship with libraries should be mentioned. One of these, collecting, preserving, and maintaining university archives is regularly administrated as a library function.

While less often within the purview of libraries, management of the many day-to-day records generated by the various campus departments and units is an important function that should be thought about in a broader administrative context.

Centers created to assist faculty and graduate students with their teaching, to facilitate student learning, and even to promote research and scholarship have been located in a number of libraries, though their reporting lines usually have been elsewhere. With the influx into the library of technologies intended to help faculty with teaching, research, and scholarship, these pedagogical and faculty support operations have some common cause with library programs. The Boston College University [18] Libraries, through the Connors Family Learning Center, promote faculty and graduate student teaching effectiveness and undergraduate learning.

2.5 Designing and Renovating New Learning Spaces

In anticipation of a project to reconfigure its library space, Washington and Lee University [19] conducted a study of library renovations and additions around the United States. The libraries examined ranged from those at liberal arts colleges and small universities to research institutions. The survey identified a number of commonalities across the construction projects.

The Washington and Lee findings echo those of others involved in recent library reconstruction. In interviews, on web sites, and in the literature again and again one

encounters the words “flexibility,” “adaptability,” “comfort,” “usability,” and “variation.” The importance of having inviting spaces that one wants to occupy is evident. To assure maximum ability to change layout and use, there should be a minimal number of fixed walls. Where feasible, soft walls—for example, made of fabric—[1] might be used. Classrooms can be designed so that they function in many modes, for larger group use, but changeable into areas for breakout sessions. Having furniture that is easily moveable is important. Library users should be permitted to configure their seating in ways useful and comfortable to them; chairs, couches, tables, and other furnishings can periodically be returned to original positions, but users should not be restricted in how they function in their work and study environment, so long as they are respectful of others. Where furniture might need to be fixed to the floor for purposes of equipment location or data or power availability, oversized surfaces ought to be provided to enable use of contiguous space by more than one person and to permit spreading out of materials. As much as possible flexible lighting and power grids are being recommended, with the display areas of large department stores, such as IKEA, seen as examples [14].

Libraries are rediscovering inter-shelving of books and other media, a practice experimented with in the 1970’s but not widely adopted then due to complexities of equipment provision and maintenance. Monographs and bound print volumes of all types, can be housed on the same shelves as CD’s, DVD’s, and other computer retrievable information storage devices, rather than having collections separately stored by format. Of course, streaming of digital content soon may make obsolete many of these storage technologies. Radio Frequency Identification capability is making easier the tasks of stack maintenance, circulation of materials, and statistical compilation.

“Each person [has] a slightly different description of the ideal workspace, ranging from a desire for total quiet and privacy to a need for the buzz of a social space...” To accommodate different desires at different times, “varied reading rooms” are suggested in *Ideas and Inspirations*, a consideration of 21st century library possibilities by New York University’s Bobst Library [20], “...a variety of public work spaces...[could affect] a range of scholars that reflect a diversity of working styles.” “The library could serve...disparate needs by establishing ‘zones,’ each one representing a different style and specific type of expected behavior,” This is precisely what the University of California, Merced has done. Floors and areas were planned as a series of different “affects,” depending upon the type of activity likely to occur in a space and how the users on the particular floor or in a particular space would function and behave. The “zones” or “affects” range from individual to collaborative, from “classic private library” to “diner study room,” from formal living room to lounging play room [21]. The university librarian there claims this approach to be a resounding success [2].

3 Conclusions

As institutions of higher education consider changes in their approaches to education and research, the role of their libraries are at the center of discussions. Some skeptics have suggested that digital developments impacting the academy would marginalize

libraries. Instead those developments, along with a millennia-old mission of providing access to information, and a newer one of information assessment, have made the clichéd “heart of the university” more vibrant than ever. Strong physical and electronic collections supportive of academic offerings and research, a physical facility that is comfortable, inviting, adaptable, and technologically sophisticated, and information and educational programming that furthers the classroom learning experience are all part of an effective academic library today.

A good physical structure includes a variety of teaching, learning, study, and research spaces, areas for individual work, group interaction, instruction, practice of presentations, and preparation of scholarly tools and products, all outfitted with sophisticated technology. Computing laboratories, electronic classrooms, training rooms, multipurpose seminar and meeting rooms, event and reception areas, and refreshment facilities all are essential. This examination of new learning spaces offers evidence that the modern academic library is maintaining its traditional mission while pursuing enhanced relationships with student and faculty support entities to the ultimate benefit of all academic clientele.

References

1. Stuart, C.: Personal communication (March 5, 2008)
2. Miller, B.: Personal communication (February 26 2008)
3. Georgia Institute of Technology Library Commons,
<http://librarycommons.gatech.edu/>
4. North Carolina State University Libraries small study rooms,
<http://www.lib.ncsu.edu/studyrooms/groupstudy.php>
5. Portland State University multimedia presentation rooms,
<http://library.pdx.edu/practicepresentationroom.html>
6. Mississippi State University multimedia presentation rooms,
<http://library.msstate.edu/services/roomreserve.asp>
7. University of Oregon Libraries smart classrooms,
<http://libweb.uoregon.edu/instruct/classrooms/>
8. North Carolina State University Libraries Digital Media Lab,
<http://www.lib.ncsu.edu/studyrooms/groupstudy.php>
9. Harvard University Lamont Library Cafeteria,
http://dining.harvard.edu/retail_dining/restaurants_lamont.html
10. Wayne State University Academic Advising Center,
<http://www.advising.wayne.edu/stsrv.php>
11. University of California, Merced Bookstore,
<http://bookstore.ucmerced.edu/StoreAbout.aspx>
12. Harvard University Lamont Library Language Resource Center,
<http://lrc.fas.harvard.edu/>
13. Florida State University Career Center,
<http://www.career.fsu.edu/library/map.html>
<http://www.career.fsu.edu/library/>
14. Meyer, R.: Personal communication (February 26, 2008)

15. University of North Carolina at Charlotte Counseling Center,
<http://library.uncc.edu/map/counselingcenter/>
16. Arizona State University Writing Center,
<http://www.asu.edu/tour/west/flh1b.html>
17. Colorado College Writing Center,
<http://www.coloradocollege.edu/learningcommons/writingcenter/>
18. Boston College Connors Family Learning Center,
<http://www.bc.edu/libraries/help/tutoring/staff.html>
19. Washington and Lee University: Survey of Library Renovation Projects A Random Sample (2006)
20. New York University Bobst Library.: NYU Bobst Library 21st Century Project Ideas & Inspirations. New York University Division of Libraries, New York (2006)
21. University of California, Merced Library Zones,
<http://ucmercedlibrary.info/about-the-library/the-story-of-the-library.html>

The Role of Information Literacy in Environmental Scanning as a Strategic Information System - A Study of Singapore SMEs

Xue Zhang, Shaheen Majid, and Schubert Foo

Division of Information Studies, 31 Nanyang Link, WKWSCI Building, Singapore 637718
{ZH0002UE, asmajid}@ntu.edu.sg, sfoo@pmail.ntu.edu.sg

Abstract. The development of information technology and telecommunication has created both opportunities and challenges for business organizations. On the one hand, it provides various channels and applications for accessing, processing and distributing information; on the other hand, it imposes higher requirements of information literacy (IL) skills of employees for dealing with information. Despite the number of studies undertaken to investigate the role of IL in the workplace, few of them have integrated IL skills with a specific business management activity, and even fewer of them have tried to evaluate the impact of IL on real business applications. This study aims to investigate the role of IL skills in environmental scanning (ES), a strategic information system used by organizations to cope with environmental changes. It will use both quantitative and qualitative methods: the quantitative approach through questionnaire survey will focus on finding statistically significant effects; the qualitative method through face-to-face interview is expected to discover more explorative information. The scope of the research is limited to SMEs (Small and Medium Enterprises), as they form a significant chunk of the economy in many countries, and they are more sensitive towards environmental uncertainties due to their limited financial resources. This study is expected to fill the knowledge gaps and build up a new model of environmental scanning as a systematic information system with consideration of the effect of IL skills and information technology applications.

Keywords: Information literacy, environmental scanning, SME, Singapore.

1 Introduction

Advancement of information and telecommunication technology has facilitated vast improvements in developing sophisticated infrastructures, which makes a huge amount of information available to people with easy and flexible access, and also provides a variety of applications and channels for processing and distributing information. However, opportunities come with challenges. Without proper skills to deal with information as well as the related technologies and applications, people suffer from various problems, such as information overload, inability to locate and extract relevant information and disorganization of information. In the workplace, employees

must possess adequate information literacy skills to search, process and evaluate information to conduct research, make decisions and solve problems, with support from information technology applications. Previous researchers have highlighted the importance of information and information literacy skills for business organizations, and they have also detected serious problems due to lack of knowledge and information handling skills. Despite the number of studies that have been conducted regarding the role of information literacy in the workplace, few of them have integrated IL skills with a specific business management activity, and even fewer of them have tried to evaluate the impact of IL skills on real business applications with consideration of the support from a suitable and effective information system. Environmental scanning, a strategic information system used by organizations to cope with environmental changes, is a typical business management process with application of both information literacy skills and information technology. This study aims to investigate the role of information literacy skills in environmental scanning for Singapore SMEs (Small and Medium Enterprises).

2 Literature Review

2.1 Definition of Information Literacy

The term “information literacy” was coined by Paul Zurkowski in the 1970s to bring attention to the needs of people working in the newly emerging technological environment [1]. Since then, the concept has been mainly used by information specialists, and promulgated worldwide through the work of the American Library Association (ALA) and the National Forum for Information Literacy [2].

However, there is no agreed definition of the term. Some researchers describe IL as requisite to lifelong learning [3], [4], while others perceive it as a natural extension of the concept of literacy in our society [5], [6]. Some have acquainted IL with information technology [7], while others have used it interchangeably with library skills [8]. The 1989 Final Report of the American Library Association’s Presidential Committee on Information Literacy, a milestone in the history of IL research, not only recognized the importance of the term, but also sought to define the skills of IL [9]. This definition was widely accepted: to be information literate, a person must be able to recognize the need for information, to effectively access, evaluate and creatively use information. Todd, Lamb and McNicholas [10] defined IL as “a holistic, interactive learning process encompassing the skills of defining, locating, selecting, organizing, presenting, and evaluating information”. Goad [11] gave a brief definition as “the ability to search for, find, evaluate, and use information from a variety of sources”.

2.2 Information Literacy in the Workplace

A number of researchers have pointed out the importance of information and IL skills in the workplace. Porter and Miller [12] report information as one of the most

important elements in competitive advantages. Forward-looking companies take the view that information is a strategic asset of the enterprise in much the same way as a company's financial resources, capital equipment and real estate, and properly employed information assets would create additional value with a measurable return on investment, and can be leveraged into strong competitiveness [13]. Drucker [14] elaborates on the need for organizations to become information literate. He suggests that corporations need to learn to ask questions as: What information do we need in this company? In what form and how do we get it? Mutch [15] also pointed out the potential importance of IL skills to business as he outlined how the concept might be employed within the business field. IL is a means of helping individuals handle the massive amount of information that pervade their daily life [16]. Karim and Hussein [13] state that good and quality information can improve decision-making, enhance efficiency and allow organizations to gain competitive advantage.

Despite its importance highlighted in the literature, information literacy, the key to information power, has not been of great concern in the business sector. Employees tend to attend more to the need for computer skills than IL skills [17]. Nevertheless, having the ability to handle technology does not necessarily mean that they are information literate [18]. Negative examples were observed in various workplace contexts, such as "unable to determine the nature and the extent of the information needed", "unable to retrieve information effectively from the information systems", "not aware of the full range of resources available" and so on, which may result in increased operating cost and inability to fully exploit valuable information sources [18], [19].

2.3 Definition and Process of Environmental Scanning

In the field of ES, the first notable study was carried out by Aguilar [20]. Aguilar defines environmental scanning as acquiring information about events and relationships in a company's outside environment, the knowledge of which would assist top management in its task of charting the company's future course of action. Subsequent studies reinforced Aguilar's definition without substantially altering this perspective; however, the process of ES was gradually extended and has been conceptualized as an integrated information management system [21], [22], [23].

Synthesizing the reviewed literature, a six-step ES process is proposed (Figure1). With clearly defined scanning needs, the formal ES process starts. Organizations actively collect environmental information through various channels and from various information sources, then process and synthesize the acquired information with the existing organizational knowledge. The processed environmental intelligence may be stored in an organization knowledge repository for future action, or disseminated to target users for final evaluation and use. If the end-user's information need is not satisfied, he may initiate a new round of ES. It is worth noting that, sometimes steps like "information processing and synthesizing" and "information distribution" may be skipped due to certain factors, such as limited time or shortage of human resources, or the collector of information will use it by him/herself without sharing it with others.

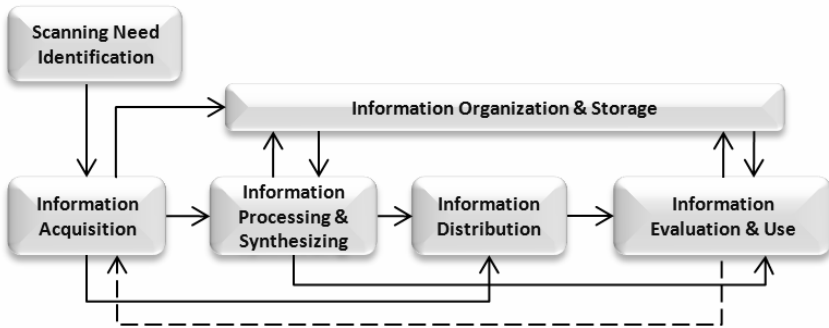


Fig. 1. Environmental Scanning Process

Although the majority of recent studies define ES as a systematic process with several steps, few empirical studies investigating ES in practice have paid equal attention to each of its steps. Most studies only focused on information collection while neglecting the rest [24], [25], [26].

2.4 Environmental Scanning as a Strategic Information System

Harrod's Librarians' Glossary [27] defines an information system as an organized procedure for collecting, processing, storing, and retrieving information to satisfy a variety of needs. Avinson and Fitzgerald [28] pointed out that an information system is a human activity system, which may or may not involve the use of a computer system. It can be either computer-based or paper-based.

The information system of an organization will be required to help it analyze the business along with its environment, formulate strategies and check whether it achieves its goals [28]. Organizations scan their business environment for market opportunities, for threats to present competitive positions, for potential replacement of currently marketed products and services, and for possible acquisitions and expansions. Such information is regarded as strategic information, and systems designed to acquire, store, organize and make available for use such information are strategic information systems [2]. From the information perspective, the ES process, also known as an information scanning system [23], is a strategic information system in the workplace.

In the literature, several studies have been done to identify the influence of an organization's information system on the ES effectiveness and hence organizational performance. For example, Subramanian, Fernandes and Harper [29] found that firms having advanced systems to monitor events in the external environment exhibited higher growth and greater profitability than firms that did not have such systems. Similarly, Ahituv, Zif, and Machlin [30] discovered significant differences in terms of the use of information systems and ES levels between more successful and less successful firms in introducing new products into the market. The differences are in

the ES pattern and frequency, the number of computerized applications, and the number of advanced marketing information systems.

2.5 Information Literacy Skills at Each Step of Environmental Scanning as a Strategic Information System

While IL is not mentioned specifically in mainstream ES literature, it is an implicit aspect of ES, as each ES step could only be completed effectively by people with the corresponding IL skills with support from an organizational information system. Specifically, employees should possess IL skills to identify information needs and to locate the best sources to obtain accurate and current information, as well as the abilities to present the information effectively to the intended audience after data mining and repackaging, which is extremely valuable to the overall effectiveness of ES and hence the success of the organization. By improving their own skills of creating, acquiring and transferring knowledge, they enable their organization to modify its behavior according to the continuously changing external environment [31]. Without employees possessing proper IL skills, organizations would not be able to conduct effective ES activities, and hence achieve alignment with the external environment.

Identification of Scanning Needs. The first step in developing an ES strategy is to accurately appraise the information needs of the various individuals and groups in the organization. Understanding their information needs and requirements would be a significant step in developing information strategy and tools for providing effective information services and promoting organization-wide creativity and innovation [13]. In the context of ES, information needs occur when the existing organizational knowledge cannot cope with the signals from the external environment. In other words, organizations need to acquire more information to analyze environmental uncertainty and adapt to it accordingly. A good recognition of scanning needs at all points in the organization is essential, as they determine the scope and depth of ES.

In an organization, the identification of information needs begins with an analysis of key decision-makers and the environment of the organization, and key decision-makers are found to be not only at the top of the organization, but also among middle managers and tactical employees [32]. In other words, identification of an organization's information needs starts from identification of individual information needs. To ensure the organization's scanning needs to be well-captured, first of all, individuals working in the organization must be able to identify their own information need clearly; secondly, they must possess essential communication skills to express their needs; finally, the person in charge of the aggregation of individual information needs should be capable in information processing and synthesizing, and conclude the organization's scanning need.

Information Acquisition. Information acquisition aims to satisfy the identified information needs. In the literature, three key issues are highlighted during the process of information acquisition: where to collect, how to collect, and when to stop.

“Where to collect” regards the source of information. Case [33] categorized information sources as internal (the company manager and staff) and external (printed and broadcast media), which is not exhaustive. Choo [34] divided information sources into three categories: textual sources, online sources and human sources. Information literate workers should realize that each kind of information source has its own advantages and disadvantages, and information sources need to be matched with the information needs and strategic objectives as well as their “accessibility” and “reliability”. For example, textual sources are well suited to situations when the information is structured and formal, or when the transmission accuracy of information is highly demanded; online sources are especially useful when reasonably complete and up-to-date information needs to be gathered swiftly; human sources tends to be preferred when dealing with ambiguous, unstructured problem situations [34].

“How to collect” concerns the methods or techniques used for gathering information. They could be routinely getting information through various media channels like newspapers, market reports or television, or acquiring first-hand data through active research methodologies like questionnaires, interviews and participant observation, or passively receiving information through subscribed alerting services provided by information vendors. With the number of methods and techniques available, people in charge of collection of environmental information should be able to select the most appropriate one, with consideration of the quality of information and the cost of collection. Moreover, information literate workers would be able to formulate a suitable search strategy which enables them to retrieve information effectively through various channels. Last but not least, collectors should be aware that the methods and techniques hired should be based on legal collection of open-source or public domain information, without involving immoral, unethical or illegal activities.

“When to stop” is about the judgment of “enough” information to satisfy the identified information needs. Both qualitative and quantitative criteria are helpful for making rational choices to determine when the collected information is “enough” [35]. The personal judgment of experienced information workers would also help identify the quantity of collection.

Information Organization and Storage. Acquired or created information should be organized and stored systematically in order to facilitate future information retrieval and sharing. In enterprises, information on paper originals could be stored in a traditional filing system, or digitized and archived on hard disks attached to file servers. No matter in which format, the design and performance of the organization and storage system, such as its creation of taxonomies, resource description and comprehensiveness, would highly affect the accessibility and retrieval of stored information, especially when the majority of the information is collected from electronic sources and the Internet. Without IL skills, organization would not be able to organize their knowledge base properly which may result in various barriers for future retrieval and use.

Information Processing and Synthesizing. The collected or generated information could be directly stored for future accessibility, or processed into information products or services through some sets of value-added activities, such as filtering, interpreting and repackaging. Analyzing the collected information and extracting meaning from it is the most important part of ES; moreover, in today's complex and turbulent environment places there is a premium on the reliability and quality of information. The collected information should be analyzed for issues and trends that may influence the organization, to assist users to acquire a better sense of situations and make better decisions, and hence facilitate the creation of a dynamic knowledge capability. The relevant information from each source should be extracted and information from multiple sources should be organized. Srinivas [36] pointed out that questions needing to be addressed during processing are: Which parts of the information collected will be used? What additional data is needed? How can information be best presented to enable situation understanding and problem-solving?

However, a recent study reports that knowledge workers are spending more time collecting information and less time analyzing it [37]. Inadequate filtering of information would result in information overload; with inadequate time for analysis; the collected information will provide either a recital of facts or a "dump" of data with little advice or confirmation [32]. Without proper information processing skills, the gathered information would be underutilized as "the organization does not know what it knows" [38].

Moreover, there are more than 100 different analytical techniques which could be used to glean meaning from the collected data and information, such as blind spot analysis, competitor benchmarking and SWOT analysis [32], and due to the rapid technological development, more advanced information systems equipped with enterprise decision support tools are available. However, these tools would still rely heavily on human interpretation and cognition [13]. If staff have insufficient knowledge of those techniques, and are without the ability to manage information flows for future utilization and development, advances in information and communication technology may also impose an immense challenge for people to handle the existing over loaded information [13].

Information Distribution. The processed environmental information, with potential effects on the organization, should be reported to the appropriate decision-makers within the firm. Myburgh [32] and Albright [22] suggest some points deserving special attention in information distribution. The first one is to ensure that the correct information or intelligence makes its way to the correct destination, as the decision-makers may be scattered throughout the organization; secondly, the information should be delivered through vehicles and in formats that mesh well with the user's information preferences and work habits; thirdly, the intelligence also must match the users' requirements of presentation, such as its orientation and content. Briefly, the real issue is getting the right information to the right person at the right time and in a usable form.

Moreover, the benefits of a wider distribution of information are also highlighted in the literature. Nutt [39], from the perspective of decision-making theory, found that when the same piece of information is distributed to many individuals, multiple interpretations could be resolved and a consensus reached. Daft [40] discovered that

multiple interpretations of the same information could improve decisions by redefining the problem. A wider distribution of information may bring more broadly based and more frequent organizational learning, as retrieval efforts are more likely to succeed and individuals and units are more likely to be able to learn [41].

Information Evaluation and Use. On receiving the processed information, the end-users would finalize it to be ready for assisting decision-making. In the current information-intensive business environment, the utilization of information is indeed a critical factor in the achievement of organizational success [42]. Information literate decision-makers would be open-minded and objective, relying not merely on the guidance of instincts and their experience. At this stage, various IL skills are required. For example, decision-makers need information evaluation skills to make judgments about the quantity and quality of the received information in terms of reliability, accuracy, timeliness and so on. If they find the information to be insufficient or unqualified, they may re-identify their scanning needs and start a new round of ES; with sufficient and high-quality information, they may still need to process and synthesize it based on the real-time situation and different usages.

3 Proposed Research Models

Based on Daft's widely accepted model [43] (Figure 2), we proposed a refined model of ES as a strategic information system (Figure 3). It is believed that perceived strategic uncertainty (PSU) would trigger the need for scanning. In Daft's model, the measurement of PSU is structured $PI \times (C+R)$, where PI =perceived sector importance; $C+R$ =perceived sector uncertainty; C =the perceived sector complexity; R =the perceived sector rate of change; the measurement of ES is scanning frequency (how often environmental information is collected) and scanning mode (use of personal or non-personal, internal or external information sources).

In our model, three factors are proposed to measure the ES process, and they are not restricted to the step of information collection. The three factors are "implemented frequency of each ES step", "rate of interest of each ES step" and "actual effectiveness of each ES step". "Rate of interest of each ES step" refers to the importance attached to each ES activity, in terms of conducting manner (primitive, ad hoc, reactive and proactive) [44] and assigned scanning unit [45]. The "implemented frequency of each ES step" and "rate of interest of each ES step" are proposed to be determined by the combination of PSU and "perceived effectiveness of each ES step". For example, with the same PSU level, higher perceived effectiveness of an ES step may result in less frequency and less interest of that ES step. With the same perceived effectiveness of an ES step, higher PSU would result in higher frequency and interest of that ES step. The "actual effectiveness of each ES step" is proposed to be determined, new from an information perspective, by the "actual effectiveness of technology and media" and "actual level of IL skills". The "actual effectiveness of each ES step", "implemented frequency of each ES step" and "rate of interest for each ES step" would together impact on the quality of the actual ES product, which would be used for assisting tactical and strategic decision-making.

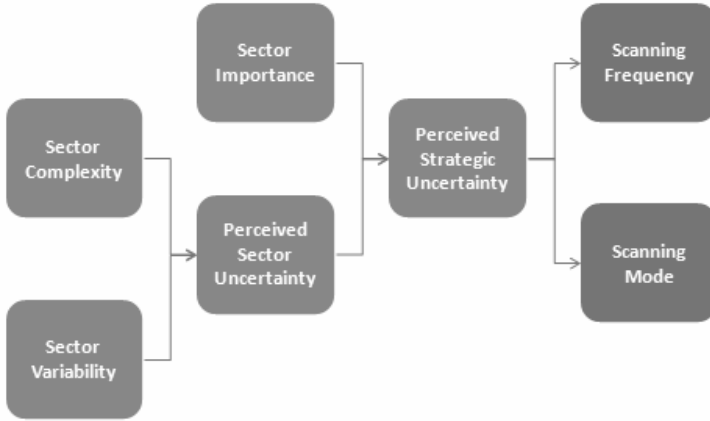


Fig. 2. PSU and ES Adapted from [43]

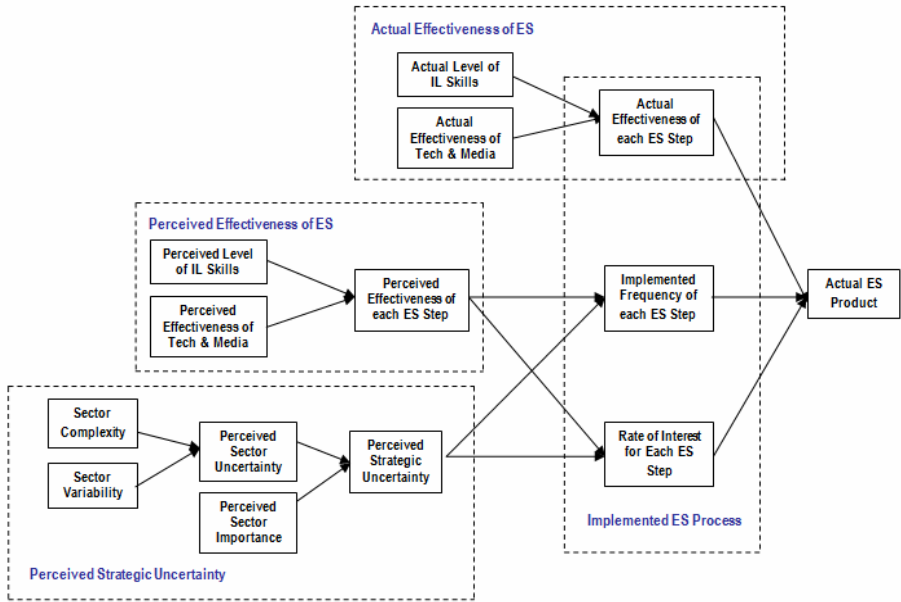


Fig. 3. PSU and Step-based ES Process

For the step of information collection, two more factors regarding the accessibility of various information sources need to be taken into consideration, which is particularly important for SMEs with limited financial investment in information technologies and application, as well as unavailability of qualified information specialists.

Specifically, the “perceived access to information sources” would help decide the frequency and rate of interest assigned to information collection, while the “actual access to information sources” would impact the effectiveness of information collection (Figure 4).

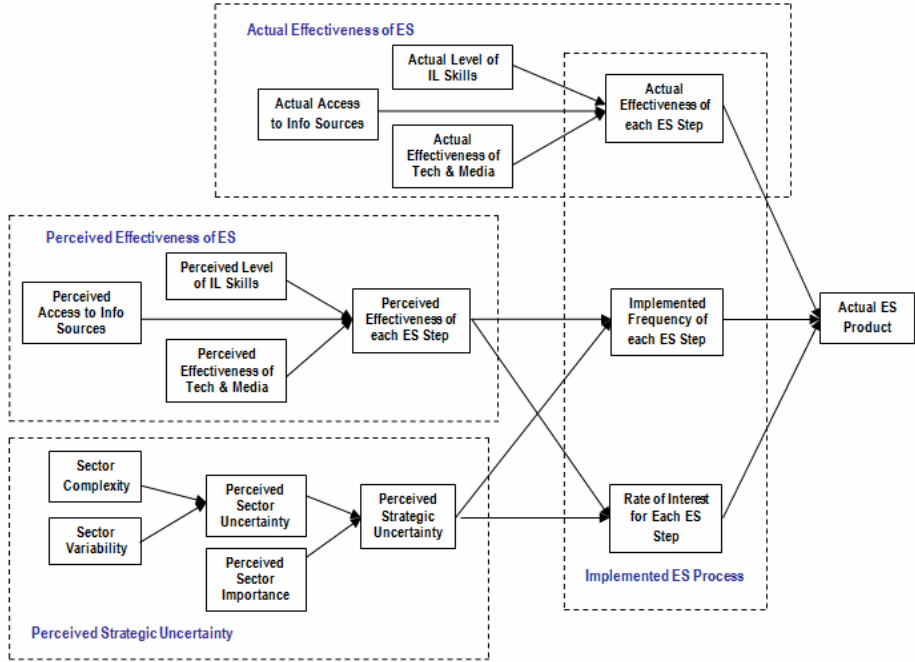


Fig. 4. PSU and ES (Information Collection)

4 Proposed Research Methodology

4.1 Research Design

A combination approach of both quantitative and qualitative method is proposed. The quantitative method through questionnaire survey/test would help obtain a general picture of the ES process under different PSU situations, statistically reveal the impact of IL skills and effectiveness of information technologies and media on the overall quality of ES products, and verify the research models proposed. The qualitative method through face-to-face interviews is expected to provide more profound information regarding ES activities and the impact of PSU, IL skills and information technologies and media, and refine the research models based on the ES participants’ perspective.

4.2 Quantitative Method through Questionnaire Survey/Test

Survey, as a relatively standardized and systematic approach of obtaining self-reported information about the beliefs, attitudes, behaviors, opinions or other characteristics of a specific population [46], [47], is chosen to acquire a general picture of the ES activities conducted by Singapore SMEs. The target major respondents are chief executive officers, who would be responsible for the overall strategy and performance of the organization. However, some sections of the questionnaire may need to be completed by people engaged in the corresponding ES activities. The survey method is selected for three main reasons:

Firstly, since there have been very few studies on the scanning behavior of Singapore SMEs, and no prior studies have integrated the role of IL skills into the ES process, this study needed to collect data to obtain a broad picture of how SMEs in Singapore scan their environment. Survey research is probably the best method available to collect original data for describing a population too large to observe directly [48]. Questionnaire survey is an economical and efficient way to maximize the coverage of the sample size as compared to other approaches.

Secondly, a major part of the study is concerned with the respondents' perceptions, such as "perceived sector importance", "perceived access to information sources" and "perceived level of IL skills". A questionnaire would enable respondents to report their perceptions while remaining anonymous and honest on the feedback.

Thirdly, the quantitative survey data could be analyzed to statistically investigate the relationship among factors proposed in the structural equation model, for example, the relationship between PSU and each step of ES in terms of frequency and rate of interest, the impact of related IL skills on the effectiveness of a specific ES step and hence the quality of the final ES products

Qualitative Method through Face-to-Face Interview. Based on the general understanding acquired from the quantitative data analysis, a face-to-face interview is proposed to ensure the gathering of more profound data from different hierarchical levels, which also allows the interviewees to express their feelings based on their own perspectives and experiences. Prior studies have found that scanning activities could be completed through staff from different functional units and at different hierarchical levels [49], [50]. However, the reviewed ES studies have focused on top management (e.g. chief executive officers, managing directors), and paid insufficient attention to middle level managers or employees, who may also play an important role in the ES process. Interviewing staff engaged in ES activities with different functional roles and from various hierarchical levels would enable the researcher to acquire more in-depth data, and increase the chance of obtaining a more reliable overall picture of ES activities. Moreover, the interview method would enable the researcher to explore the ES activities conducted by the same company in different time periods through recall of various strategic situations.

Proposed Sample Selection and Characteristics. For better concentration, only two industries are proposed to be included in this study, i.e. travel agents and food manufacturers. These two industries are all dominated by small and medium sized enterprises (SMEs), but they are from different consumption layers. The food manufacturing industry is comparatively more stable as the products are kinds of human necessities, and companies' operation may not be sensitive to environmental changes. However, travel agencies operate in a more dynamic environment. They need to react immediately to environmental changes and their performance is determined by the economic conditions of the local economy.

According to NATAS (National Association of Travel Agents Singapore), there are 331 active members [51]; based on the directory provided by SFMA (Singapore Food Manufacturers' Association), there are 297 food manufacturers [52] The sample for this study would be selected from the companies fulfilling two criteria. First, company size should be above 20 employees, which ensures sufficient manpower for conducting systematic ES activities. Second, those companies must be concentrating on a single business domain without operating across different industries. This is to ensure that the firms would focus on the task environment of their primary business without paying attention to multiple operating environments as occurs in diversified firms.

5 Significance of the Study

5.1 From a Theoretical Perspective

As mentioned in the literature review, besides information acquisition, insufficient attention has been paid to the other steps of ES. Moreover, IL and information technology applications, as enablers to conduct effective ES activities, have not been highlighted in the reviewed studies. This research aims to address these problems and the findings are expected to fill in the knowledge gaps and build up a new model of ES as a systematic process with consideration of the effect of IL and information technology applications.

5.2 From a Practical Perspective

The findings of this study, to some extent, can provide insights into the current situation of SMEs in Singapore regarding their use of IL skills and information technology for ES, reveal their ES mechanisms, detect the problems they may encounter during the scanning process, and generate implications on how to improve the current situation. The study findings are expected to reveal the contribution of IL skills and adoption of information technology applications to conducting more effective ES activities, the results of which would help decision-makers to do better strategic planning, to achieve strategy-environment alignment, and hence contribute to the organizational performance. These demonstrated links would be able to create awareness of the power of information among SMEs, the importance of IL skills and adoption of a suitable

information system and applications, as well as the significance of conducting ES. In the long term, SMEs could become information-literate learning organizations, responding to the changes and new threats from the environment in a timely manner, and thus survive and succeed in both local and international markets.

5.3 From a Methodological Perspective

The majority of previous studies have investigated ES activities using quantitative methods such as questionnaire survey or survey-based interview, even those trying to demonstrate the contribution of ES to organizational performance. Only limited qualitative studies have used interviews, and these interviews were limited to the top management level. This study will explore the feasibility of adopting both quantitative and qualitative methods to generate a quality set of findings and to cross validate them.

6 Conclusion

ES could provide early warning signals for organizations, and help companies develop and modify business strategies to meet changing external circumstances and hence improve their competitiveness and performance. In today's turbulent environment, to conduct effective ES activities, employees must possess the corresponding IL skills with support from the various information technology applications.

In the reviewed literature, the majority of studies investigating ES activities have mainly focused on the information collection step, and insufficient attention has been paid to the role of IL skills and information technology applications. For methodology, only limited interviews were conducted at the top management level. The common limitations of prior studies have been addressed in the proposed model through the equal attention paid to each ES step and the newly added variables as "perceived/actual level of IL skills" and "perceived/actual effectiveness of technology and media". Moreover, our proposed research design, as a combination of both quantitative and qualitative methods, expects to enrich this set of information by obtaining inputs across the hierarchy of the organization, thereby providing a richer and more accurate picture of ES by the various stakeholders involved in this process.

References

1. Kapitzke, C.: Information Literacy: A Positivist Epistemology and a Politics of Outformation. *Educational Theory* 53, 37–53 (2003)
2. Feather, J., Sturges, P. (eds.): *International Encyclopedia of Library and Information Science* (2003)
3. Hancock, V.E.: *Information Literacy for Lifelong Learning*. ERIC Clearinghouse on Information & Technology, Syracuse University, N.Y. (1993)
4. Moore, P.: *An Analysis of Information Literacy Education Worldwide*. White Paper prepared for UNESCO, the U.S. National Commission on Libraries and Information Science, and the National Forum on Information Literacy, for use at the Information Literacy Meeting of Experts (2002)
5. Bruce, C.: *Information Literacy as a Catalyst for Educational Change: A Background Paper*. Information Literacy Meeting of Experts (2002)

6. Stern, C.M.: Information Literacy “unplugged”: Teaching Information Literacy Worldwide Without Technology. White Paper prepared for UNESCO, the U.S. National Commission on Libraries and Information Science, and the National Forum on Information Literacy. Information Literacy Meeting of Experts, Prague, The Czech Republic (2002)
7. Mitchell, P.: Online Technology in the Curriculum: A Reality. ASLA XIV Beyond the Horizon Fremantle, Australia (1996)
8. Kuhlthau, C.C.: Information Skills for an Information Society: A Review Of Research. Information Reports and Bibliographies 19, 14–26 (1990)
9. ALA: A Progress Report On Information Literacy: An Update On The American Library Association Presidential Committee On Information Literacy: Final Report (1989)
10. Todd, R.J., Lamb, L., McNicholas, C.: The Power of Information Literacy: Unity of Education And Resources for the 21st Century. In: Annual Meeting of the International Association of School Librarianship (1992)
11. Goad, T.W.: Information Literacy and Workplace Performance. Quorum Books, Westport (2002)
12. Porter, M.E., Miller, V.E.: How Information Gives Competitive Advantages. Harvard Business Review 63, 110–128 (1985)
13. Karim, N.S.A., Hussein, R.: Managers’ Perception of Information Management and the Role of Information and Knowledge Managers: The Malaysian Perspectives. International Journal of Information Management 28, 114–127 (2008)
14. Drucker, P.: Be Data Literate - Know What To Know. The Wall Street Journal 16 (1992)
15. Mutch, A.: Information Literacy: An Exploration. International Journal of Information Management 17, 377–386 (1997)
16. Lyman, P.: Information Literacy. Liberal Education 87, 28–37 (2001)
17. Bruce, C.: Workplace Experiences of Information Literacy. International Journal of Information Management 19, 33–47 (1999)
18. Cheuk, B.: Information Literacy in the Workplace Context: Issues, Best Practices And Challenges. Information Literacy Meeting of Experts (2002)
19. O’Sullivan, C.: Is Information Literacy Relevant in the Real World? Reference Services Review 30, 7–14 (2002)
20. Aguilar, F.J.: Scanning the Business Environment. Macmillan, New York (1967)
21. Aaker, D.A.: Organizing a Strategic Information Scanning System. California Management Review 25, 76–83 (1983)
22. Albright, K.S.: Environmental Scanning: Radar for Success. Information Management Journal 38, 38–45 (2004)
23. Costa, J.: An Empirical-Based Review of The Concept of Environmental Scanning. International Journal of Contemporary Hospitality Management 7, 4–9 (1995)
24. Ebrahimi, B.P.: Perceived Strategic Uncertainty and Environmental Scanning Behavior of Hong Kong Chinese Executives. Journal of Business Research 49, 67–77 (2000)
25. McGee, J.E., Sawyerr, O.O.: Uncertainty and Information Search Activities: A Study of Owner-Managers of Small High-Technology Manufacturing Firms. Journal of Small Business Management 41, 385–401 (2003)
26. Strandholm, K., Kumar, K.: Differences in Environmental Scanning Activities Between Large and Small Organizations: The Advantage Of Size. Journal of American Academy of Business 3, 416–421 (2003)
27. Prytherch, R. (ed.): Harrod’s Librarians’ Glossary: 9,000 Terms Used in Information Management, Library Science, Publishing, The Book Trades, And Archive Management. Gower Publishing Company, England (1995)
28. Avinson, D.E., Fitzgerald, G.: Information systems development: methodologies, techniques and tools. McGraw-Hill, New York (2006)

29. Subramanian, R., Fernandes, N., Harper, E.: Environmental Scanning in U.S. Companies: Their Nature And Their Relationship To Performance. *Management International Review* 33, 271–286 (1993)
30. Ahituv, N., Zif, J., Machlin, I.: Environmental Scanning And Information Systems In Relation To Success In Introducing New Products. *Information and Management* 33, 201–211 (1998)
31. Giesecke, J., McNeil, B.: Transitioning to the Learning Organization. *Library Trends* 53, 54–67 (2004)
32. Myburgh, S.: Competitive Intelligence: Bridging Organizational Boundaries. *Information Management Journal* 38, 46–55 (2004)
33. Case, D.O.: *Looking for Information: A Survey of Research on Information Seeking, Needs, And Behavior*. Academic Press, New York (2002)
34. Choo, C.W.: *Information Management for the Intelligent Organization, the Art of Scanning the Environment*. Information Today, Inc., Medford (2002)
35. Prabha, C., Connaway, L.S.: What is Enough? Satisfying Information Needs. *Journal of Documentation* 63, 74–89 (2007)
36. Srinivas, H.: *Information for Decision-Making* (2009), <http://www.gdrc.org/decision/info-decision.html>
37. Anonymous: Knowledge Workers: Shifting Research Information Sources. *T + D* 59, 15 (2005)
38. O'Dell, C., Grayson, C.J.: If Only we Knew What we Know: Identification And Transfer Of Internal Best Practices. *California Management Review* 40, 154–174 (1998)
39. Nutt, P.C.: Surprising But True: Half The Decisions In Organizations Fail. *Academy of Management Executive* 13, 75–90 (1999)
40. Daft, R.L.: *Essentials of Organizational Theory and Design*. South-Western College Publication, Cincinnati (2001)
41. Huber, G.P.: Organizational Learning: The Contributing Processes and the Literature. *Organization Science*. A Journal of the Institute of Management Science 2, 88–115 (1991)
42. Souchon, A.L., Cadogan, J.W., Procter, D.B., Dewsnap, B.: Marketing Information Use and Organizational Performance: The Mediating Role of Responsiveness. *Journal of Strategic Marketing* 12, 231–242 (2004)
43. Daft, R.L., Sormunen, J., Parks, D.: Chief Executive Scanning, Environmental Characteristics, and Company Performance. *Strategic Management Journal* 9, 123–139 (1988)
44. Jain, S.C.: Environmental Scanning in U.S. Corporations. *Long Range Planning* 17, 117–128 (1984)
45. Olsen, M.D., Murthy, B., Teare, R.: CEO Perspectives on Scanning the Global Hotel Business Environment. *International Journal of Contemporary Hospitality Management* 6, 3–9 (1994)
46. Edward, J.E., Thomas, M.D., Rosenfeld, P., Booth-Kewley, S.: *How to Conduct Organizational Surveys: A Step-By-Step Guide*. Sage Publications, Thousand Oaks (1997)
47. Fowler, F.J.: *Survey Research Methods*. Thousand Oak, California (2002)
48. Babbie, E.R.: *Survey Research Methods*. Wadsworth Publishing, Belmont (1990)
49. Majid, S., Kowtha, R.: Utilizing Environmental Knowledge for Competitive Advantage. In: *International Conference on Information Resources Management* (2008)
50. Zhang, X., Majid, S.: Environmental Scanning Initiatives of SMEs in Singapore. *Libri: International Journal of Libraries and Information Services* 59, 114–123 (2009)
51. NATAS Members Directory, <http://www.natas.travelMembersDirectory.aspx>
52. SFMA Members Directory, <http://www.sfma.org.sg/tradecat.aspx?id=>

Using Internet Services for Personal Information Management

Shaheen Majid, Myint Malar San, Saw Tin Naung Tun, and Thu Zar

Nanyang Technological University, School of Communication & Information,
31 Nanyang Link, Singapore 637718
asmajid@ntu.edu.sg

Abstract. The popularity and improvement in features provided by various Internet services are leading us to a situation where personal information items can be stored and managed online. Personal Information Management (PIM) refers to a set of activities a person performs in order to acquire or create, store, organize, maintain, retrieve, use and distribute personal information for different purposes, including preserving personal and family information assets. The main purpose of this study was to investigate perceptions of university students of Internet services for managing their personal information; the type of information stored by them; how they organize, manage, access and control information distribution; and their concerns regarding privacy and security of their personal information. A pre-tested questionnaire was used for data collection and 212 students from two public universities took part in this study. It was found that 75% of the respondents were using Internet services for storing and managing some of their personal information items. The Internet services were mainly used by the students for storing email addresses and messages, personal text documents and photos. Only a small number of the students were using online storage for maintaining information about their appointments, telephone numbers, draft documents, audio and video recordings, and list of thing to be done. Over two-thirds of the respondents revealed that they only store non-sensitive personal information due to their concerns related to privacy and security of personal information. The paper also suggests certain measures for the improvement of personal information management skills of students and the general public.

Keywords: Personal information management, internet services, social networking services, information fragmentation, Singapore.

1 Introduction

Personal Information Management (PIM) is not a new concept and all individuals in their lives handle a considerable amount of information for undertaking personal responsibilities and performing diverse tasks and activities. For this purpose, they need to use different types of personal information items such as records of personal property and other assets, academic and professional transcripts, medical records, insurance policies, mortgage contracts, bank statements and records of other financial

transitions, legal documents, family photos, contact information of friends and relatives, etc. Proper management and access to this information can make a big difference in their lives. Mismanagement of personal information on the other hand, may result in loss of family heritage, delays in legal proceedings, financial losses, delays in medical treatment, and problems in making insurance and other claims. In many situations, inadequate and inappropriate management of personal information may also cause stress, anxiety and embarrassment. Indiscriminate acquisition of information and its mismanagement can also create information overload. As a result, it is now becoming crucial for all individuals in the society, whether employed or unemployed, even for housewives, senior citizens and children, to properly manage their personal information. Another related issue is that due to technological advancements, personal information may be available in many different formats such as digital documents, audio-visual recordings, and multimedia objects [1]. Many individuals may not be able to effectively manage their valuable personal information either due to inadequate awareness or lack of desired information handling competencies.

These days, due to technological advancements and popularity of digital tools and gadgets, even ordinary citizens generate voluminous amounts of electronic information artifacts. As the number of digital objects grows, people may feel the pressure to properly store and organize these items for fast and accurate retrieval. As an effort to overcome problems associated with managing personal information items, the concept of PIM is gaining popularity. Personal Information Management includes activities such as storage, organization, and retrieval of information by an individual for his/her own use [1]. A more comprehensive definition of PIM is provided by Jones [2] and it states:

Personal Information Management (PIM) refers to both the practice and the study of the activities a person performs in order to acquire or create, store, organize, maintain, retrieve, use and distribute the information needed to meet life's many goals (everyday and long-term, work-related and not) and to fulfill life's many roles and responsibilities (as parent, spouse, friend, employee, member of community, etc.). PIM places special emphasis on the organization and maintenance of personal information collections in which information items, such as paper documents, electronic documents, email messages, web references, handwritten notes, etc., are stored for later use and repeated re-use.

Jones [2] divides personal information management into seven interrelated activities which are finding, keeping, organizing, maintaining, managing information flow, measuring and evaluating, and making sense. In order to facilitate personal information management activities, a wide range of tools and systems have been developed. For example, personal digital assistants (PDAs) usually provide many PIM features such as organizer, calendar, to-do list, tasks management, notes, etc. A previous study identified 27 artifacts used by mobile workers to manage their information [3]. The 10 most commonly used artifacts were diary, mobile phone, loose paper, work file, filofax, Dictaphone, electronic PIM, address book, notepad, and laptop. Since then some of these tools have been replaced with new and more powerful tools.

The choice of tools for managing personal information is usually determined by the type of information objects to be managed. For example, a contact phone number list may be best managed by using the phone itself while personal photos may be managed and shared online by using services such as “Flickr”. The most frequently stored management personal electronic information includes telephone numbers, names and addresses, appointments, reference documents, meeting notes, rough work, to-do lists, reminders, noteworthy dates, and identification information [3]. This list is not exhaustive, as now many new types of information objects are being created and managed by people. Although different tools and gadgets can be used for effective personal information management, many people may not be aware of their features or how to effectively use them. Similarly, the ways people prefer to manage their personal information items may differ from one to another. For example, it was reported that more people prefer “organization” to “finding” as only 1 in 20 people would prefer seeking the needed information to organizing it into different folders for future retrieval [4]. In order to further improve personal information management, efforts are being made to introduce new methods and techniques such as weaving threads of actions using context [5], adaptive semantic approach [6] and associative personal information management [7].

One problem associated with using multiple tools is fragmentation of information across several devices which may be more useful in handling a specific type of information objects such as documents, emails, instant messages, photos, audio and sound recordings, etc. Even information related to an activity, e.g. a research project, may be scattered into many folders or applications (Excel, MS Word, PowerPoint, SPSS), or devices. This situation is further worsened if documents have many versions, stored at different places. Another related problem associated with personal information management is the inability of mobile devices to exchange data among them. This fragmentation may result in wastage of time in locating the needed information from multiple locations or devices or even losing it all together. Even though currently no perfect solution is available to avoid information fragmentation, this problem may become less crucial due to likely convergence of different technologies [2]. However, Barreau [8] feels that the real challenge in managing personal information is not the unavailability of appropriate technological tools but rather the lack of basic information organization skills among the general public.

Many Internet services, such as Gmail, Yahoo, Hotmail, Google Apps, Office online, Facebook, and MySpace, in addition to their respective functionalities, also provide storage space to their account holders for managing their personal information items [9]. A recent study on the use of Internet services has shown that over 54% of the users of these services belonged to the age group 16-24 years [10]. It is, therefore, worth studying whether or not young adults are also using the Internet services for managing their personal information. The main objective of this study was to understand the usage of different Internet services by the students for managing their personal information objects. Other areas covered by this study include: the reasons for using Internet services for PIM, types of information objects managed online, mechanisms used for accessing and retrieving personal information items, and finally concerns of the students regarding the privacy and security of their personal information objects.

2 Method

A pretested questionnaire was used for collecting data for this study. There were 5 sections in the questionnaire. The first section collected information about the possible reasons for using the Internet services for managing personal digital information objects. The second section was on the types and frequency of personal information stored. The third section solicited information about the mechanisms used by the respondents for organizing their personal digital items through different features provided by the Internet services. Next, the respondents were asked to indicate how the stored information objects were retrieved. Finally, the participants were asked about their perceptions of privacy and security of personal information stored using the Internet services.

Data was collected from two public universities in Singapore, i.e. Nanyang Technological University and National University of Singapore, in October, 2009. As this study focused on management of digital information objects, students from IT related disciplines (Digital Media, Signal Processing, Computer Engineering, and Computer Aided Manufacturing) were included in this survey. A convenience sampling technique was used and instructors of different courses were approached to seek their approval for conducting this survey in their respective classes during the class break time. A total of 270 questionnaires were distributed and 212 filled-in questionnaires were returned, resulting in a response rate of 78.5%.

3 Findings

The following sections provide key findings of this survey.

3.1 Demographic Information

Fifty-two percent of the participants were male and the remaining 48% were females. A majority (78%) of the respondents were in the age group of 21 to 30 years while 14% were less than 21 years old. The remaining 9% were more than 30 years old.

As both the participating universities enroll international students, a wide diversity was found in the nationality of the participating students. The majority of the students were from Singapore, China, Myanmar and India (Figure 1). A similar diversity was found in the academic programs attended by the respondents: 33% of the students came from computer engineering, 31% from signal processing, 24% from digital media, and the remaining 12% were from computer aided manufacturing.

3.2 Reasons of Using Internet Services for PIM

The respondents were asked about the Internet services used by them for storing and managing their personal digital information objects. Two-thirds of the respondents reported using such services for managing their personal information (Table 1). Those

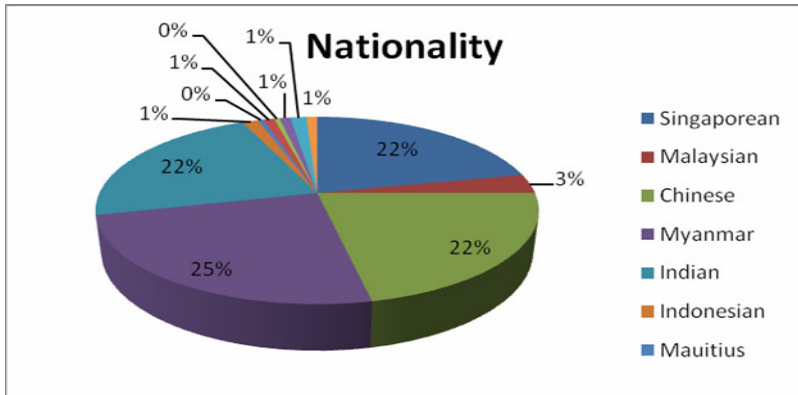


Fig. 1. Nationalities of the respondents

Table 1. Reasons for Using Internet Services for Storing Personal Information Items (N=160)

Statement	Strongly Disagree	Disagree	Neutral	Agree
Internet services make my personal digital information items available anywhere anytime.	4.4%	3.2%	8.1%	52.5%
Most internet services offer enough storage capacity for storing my personal digital information items.	3.1%	12.5%	13.1%	51.3%
Using an internet service as central storage reduces the need to maintain multiple copies of my items.	2.5%	16.9%	18.8%	50.0%
Most internet services are more stable than my own PC in terms of system stability.	5.0%	23.8%	25.0%	36.8%
Uploading and downloading speed of personal information items from internet services is satisfactory.	5.6%	20.6%	19.4%	46.9%
Personal information items can easily and effectively be organized by using features of the internet services.	2.5%	8.8%	20.6%	61.8%
Search features of Internet services allow me to retrieve my information item quickly and accurately.	0.0%	15.0%	18.8%	54.4%
Personal digital information items can easily be shared with others.	0.6%	5.7%	10.0%	65.0%
Most internet services are trust-worthy for privacy and confidential information.	4.4%	28.2%	24.3%	38.8%
Internet services are only suitable for non-sensitive information.	4.4%	12.5%	15.6%	41.9%

respondents who were using the Internet services for storing and managing their personal items were asked to indicate reasons for using these services. Over 83% of the respondents said they use Internet services as the stored information can be accessed anytime anywhere and these items can be easily shared with others. Similarly over 70% of the students agreed that the Internet services provide adequate storage space, and offer useful features for organizing and retrieving personal information items.

Table 2. Types of Personal Information Objects Stored in Space Provided by Internet Services

Digital Information Item	Use Frequency of Internet Services		
	N	Mean Score (scale: 1-5)	Standard Deviation
Email Addresses	160	3.94	1.13
Website Links	159	3.62	1.08
Text Documents	160	3.57	1.12
Mailing Addresses	158	3.48	1.28
Photos	160	3.43	1.06
Draft works - work documents for continuation at home or elsewhere	158	2.70	1.10
Sound and video recordings	160	2.67	1.20
Schedules	160	2.55	1.12
Reminders	159	2.52	1.11
Noteworthy dates	160	2.50	1.10
Videos	156	2.49	1.12
Appointments	160	2.43	1.12
To Do List	160	2.40	1.07
Telephone Numbers	160	2.18	1.07
Personal Identification Information (e.g. passport, IDs and passwords, bank accounts, etc.)	160	2.04	1.23

Another reason given for using Internet services was lack of need to maintain multiple copies of personal items (61.8%). However, only less than half of the respondents agreed that the Internet services providing personal storage space are trustworthy for maintaining private and confidential personal information.

It appeared that a sizeable majority of the students were convinced of the advantages of using Internet services for storing their personal digital items but at the same time had some concerns regarding privacy and confidentiality of their personal information. Probably that is the reason why they thought these services were not as reliable as their personal computers for storing personal confidential information.

3.3 Types of Personal Information Objects Stored

A scale of 1 to 5 was used to elicit information about the frequency at which students put their personal information in the storage provided by different online services (Table 2). As expected, email addresses were the most frequently stored information (mean score 3.94) as it is a standard feature provided by almost all email services. The next frequently stored information was bookmarks of important websites. Text documents (mean score 3.57), mailing addresses (mean score 3.48), and personal photos (mean score 3.43) were the next three frequently stored items by the students in the space provided by online services. Sound and video recordings were stored less frequently, probably due to their big file size and slow uploading and downloading speed.

Table 3. Reasons for Using Internet Services for Storing Personal Information Items (N=160)

Statement	Frequency			
	Never	Seldom	Some-times	Most of the time
I pay adequate attention to organizing my digital information items for their easy retrieval.	11.9%	18.8%	38.1%	21.9%
I use hierarchical folders for organizing my digital information items online.	7.5%	27.5%	29.4%	25.0%
I use tagging/labeling to organize my digital information items online.	7.5%	23.1%	36.9%	24.4%
I use a combination of hierarchical folder structure and tagging/labeling for organizing my digital items.	10.6%	21.9%	38.4%	22.5%
I use meaningful file and folder names which adequately represent the content of those files and folders.	6.3%	11.9%	36.8%	27.5%
I use meaningful tagging/labeling descriptions which adequately represent the content of my items.	6.3%	15.0%	29.4%	38.1%
Available features of the Internet services are useful in organizing my information items online.	5.6%	13.8%	38.1%	31.9%

Comparatively less frequently stored information items in online storage were appointments (mean score 2.43), 'to-do-lists' (mean score 2.40), and telephone numbers (mean score 2.18). Probably the respondents preferred using other tools such as Microsoft Outlook or telephone/cell phone directories for storing such information for easy access and use. The least frequently stored information items were personal identification information such as IDs and passwords of various online services, information about bank account(s), passport number, airlines' frequent flyer accounts, etc. This is understandable as previously over two-thirds of the students said that they do not keep confidential and sensitive information in the storage space provided by different Internet services.

3.4 Organizing Personal Digital Information Items Online

It is important that digital items stored in the online storage should be properly organized for their fast, accurate and convenient retrieval. A group of statements were provided to the participating students to understand their information organization behavior (Table 3). It was interesting to note that 31.2% of the students said that they either 'always' or 'most of the time' pay adequate attention for organizing their personal information items online, while almost the same number of students (30.7%) revealed that they 'seldom' or 'never' organize their information online. A similar trend was observed for the next three statements related to using hierarchical folders, tagging or labeling of information items, and the use of a combination of hierarchical folders and tagging, where around 30% of the students each were either using these features 'always/most of the time' or 'seldom/never'.

It was encouraging to note that over 45% of the students said that they use meaningful file/folder names and tag/label descriptions to adequately represent the contents of their information items. It was also worth noting that 42.5% of the students felt that the features provided by online services were useful in organizing their personal information items. On the whole, it was observed that less than half of the students were properly organizing their personal information items in the storage provided by the Internet services. An implication of this behavior is that they may face problems in retrieving their stored information items quickly and accurately, particularly once the number of online objects has grown considerably. In a worst case scenario they may lose their valuable personal information all together. There is a need to create awareness among the students about the importance of proper information organization, no matter whether stored in their personal computers or in the storage provided by online services.

3.5 Retrieval of Information Stored Online

A group of statements were used to understand information retrieval behavior of the students. Nearly 34% of the students revealed that they 'always' or 'most of the time' browse through their online folders to reach the desired item (Table 4). Another 41.9% of the students said that, if they do not know the exact location of their items, they frequently use search features to retrieve such items. Some 26.3% of the students also revealed that, due to searching convenience, they 'always' or 'most of the time' search their online items even when they know their locations. However, it was worth noting that more than 40% of the students were only 'sometimes' using various retrieval approaches, which indicates no clear preference for a particular approach. It is equally worth noting that one-half of the students accepted that 'sometimes' they encounter problems in retrieving the stored items. It could be due to their inappropriate information organization or lack of adequate browsing and searching skills or a combination of these two factors.

Table 4. Retrieval of Personal Information Items (N=160)

Statement	Frequency			
	Never	Seldom	Some-times	Most of the time
I browse through my online files and folders to reach my desired digital items	6.9%	16.8%	42.5%	26.3%
I use free text search or other search facilities to access an item if I don't know its exact location.	7.5%	15.0%	35.6%	33.1%
I use free text search or other search facilities to access an item for ease of use even if I know its location.	6.9%	23.7%	43.1%	23.8%
I can know the location of my desired items through file and folder names or label descriptions.	3.8%	15.0%	43.1%	27.5%
I find difficulty in finding the information items I need.	7.5%	32.5%	50.0%	9.4%

3.6 Perceptions of Privacy and Security Related Issues

A major concern of storing personal information items in the space provided by the Internet services was maintenance of their privacy and security. Nearly 72% of the students 'agreed' or 'strongly agreed' that they do not store their sensitive personal information, or they only store those items that they intend to share with others (Table 5). Similarly, only 31.9% of the students 'agreed' or 'strongly agreed' that online storage services can adequately protect their personal digital items. Finally, over 56.2% of the students agreed that they can control sharing of their personal digital items with the intended individuals. On the whole, it appeared that many students had reservations about the level of privacy and security provided by the Internet services; nevertheless they agreed that these services provide a useful platform for sharing personal digital items with the intended users.

Table 5. Perception of Privacy and Security of Personal Digital Items (N=212)

Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I don't put sensitive personal information in the storage provided by the internet services.	5.0%	7.50%	15.63%	31.9%	40.0%
I store only those personal information items that I want to share with others.	2.5%	10.7%	14.38%	45.0%	27.5%
Internet services can adequately protect my information items from frauds, vandalism, break-ins, theft, etc.	8.1%	35.6%	24.38%	28.8%	3.1%
I can control the level of sharing personal information items to my intended audience.	1.9%	13.1%	28.75%	50.6%	5.6%

4 Conclusion

Storage space provided by the Internet services can help account holders to put their personal digital items at a centralized location thus reducing the problem of information fragmentation. However, many potential users have reservations about the privacy and security of their personal digital items. These concerns have been exacerbated by recent vandalism attempts of highly secured websites. They feel if people can break into the highly secured websites, it will be rather easier for them to steal personal digital items. In order to encourage the use of online storage services for managing personal digital items, it is desirable that additional security measures be implemented. This study also revealed that although a majority of the students were using online storage for maintaining selected personal digital items, they were not adequately using the information organization features of these services. These days even an ordinary person produces or collects a large number of digital items; therefore, it is necessary for him/her to possess adequate knowledge of managing personal information. Libraries can take the initiative and make PIM part of user education. Once a person is able to properly manage his personal information, he can easily

apply these skills for managing his work related information. Finally if all employees in an organization know how to manage their work-related information, the organization as a whole will benefit by taking full advantage of its information and knowledge assets.

References

1. Bergman, O., Boardman, R., Gwizdka, J., Jones, W.: Personal Information Management. In: Proceedings of the International Conference Extended Abstracts on Human Factors in Computing Systems, pp. 1598–1599. ACM Press, New York (2004)
2. Jones, W.: Keeping Found Things Found: The Study and Practice of Personal Information Management. Morgan Kaufmann, San Francisco (2008)
3. Lees, D.Y., Meech, J.F., Thomas, P.J.: Information, Artifacts and Management Strategies: The Personal Perspective. In: Proceedings of Sixth Australian Conference, pp. 306–307 (1996)
4. Teevan, J., Jones, W., Capra, R.: Personal Information Management (PIM). In: Proceedings of SIGIR Forum, vol. 42, pp. 96–102. ACM Press, New York (2008)
5. Shimakawa, H., Kaede, S., Nakamura, E., Azuma, S.: Weaving Threads of Actions for Personal Information Management. In: Proceedings of IEEE International Symposium on Communication and Information Technology, pp. 787–792 (2004)
6. Kratkiewicz, G., Mitchell, G.: An Adaptive Semantic Approach to Personal Information Management. In: Proceedings of IEEE International Conference on Systems, Man and Cybernetics, vol. 2, pp. 1395–1400 (2004)
7. Diehl, J.: Associative Personal Information Management. In: Proceedings of the 27th International Conference Extended Abstracts on Human Factors in Computing Systems, pp. 3101–3104. ACM Press, Boston (2009)
8. Barreau, D.: The Persistence of Behavior and Form in the Organization of Personal Information. *Journal of the American Society for Information Science & Technology* 59, 307–317 (2008)
9. Jiang, T., He, D.: Redefining Social Network Services: A Solution to Personal Information and Knowledge Management. In: Proceeding of IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology Workshops, pp. 292–295 (2007)
10. Ofcom,
[http://www.ofcom.org.uk/advice/media_literacy/medlitpub/
 medlitpubrssi/socialnetworking/report.pdf](http://www.ofcom.org.uk/advice/media_literacy/medlitpub/medlitpubrssi/socialnetworking/report.pdf)

New Approach for Automated Categorizing and Finding Similarities in Online Persian News

Naser Ezzati Jivan¹, Mahlagha Fazeli², and Khadije Sadat Yousefi²

¹ National Library and Archives of the I.R of Iran

Ezzati@nlai.ir

² Iranian University of Science and Technology

{Mfazeli, Khyousefy}@comp.iust.ac.ir

Abstract. The Web is a great source of information where data are stored in different formats, e.g., web-pages, archive files and images. Algorithms and tools which automatically categorize web-pages have wide applications in real-life situations. A web-site which collects news from different sources can be an example of such situations. In this paper, an algorithm for categorizing news is proposed. The proposed approach is specialized to work with documents (news) written in the Persian language but it can be easily generalized to work with documents in other languages, too. There is no standard test-bench or measure to evaluate the performance of this kind of algorithms as the amount of similarity between two documents (news) is not well-defined. To test the performance of the proposed algorithm, we implemented a web-site which uses the proposed approach to find similar news. Some of the similar news items found by the algorithm has been reported.

Keywords: Categorization of web pages, category, automatic categorization of Persian news, feature, similarity, clustering, structure of web pages.

1 Introduction

Taking into account the large bulk and wide variety of web data, organizing these data for easy access and improving the search results is vital. Many efforts have been made to categorize web pages. As the web includes many different kinds of data (such as texts, images, multimedia data, etc.), there are different categorization methods for each.

These methods include: categorization of texts based on statistical and algorithmic methods of machine learning [1]. In machine learning algorithms, training data are used to train categorizers. The software used for categorizing news is called categorizer. Categorizers can categorize new pages after they have been trained. These methods include k-Nearest Neighbor approach [2], Bayesian probability models [3] [4] [5], inductive learning rules [6], backup machines [7], neural networks [8], and decision making trees [3].

The possibility of reading online news is one of the web facilities used by many users. There are a lot of sites which include daily news. If a user wants to read more

about a piece of news from other sites, he or she will have to search different sites to find similar news. This can be time-consuming for the user. To solve the problem, some methods have been proposed for categorization of news on the internet.

In this project, news headlines and summaries were used for categorization of news. Keywords for each piece of news are extracted from the headline and summaries, which are then used to collect similar news from a news data bank using a web crawler. The first part of this paper examines methods of web page categorization and in the second part approach used in the paper for categorizing Persian news is elaborated on. In the third part, the results of trying the approach out for Persian news are presented.

2 Categorization of Web Pages

Many efforts with differing degrees of precision have been made for categorization of web pages, the most important of which include:

- Manual categorizations by experts
- Cluster methods
- Content analysis of links and documents.

2.1 Manual Categorization

In the first method, some experts in each field analyze the contents of web pages and put them into different categories according to their topics. A good example of this method of categorization is that of dmoz.org which categorizes web pages using experts around the world. Yahoo had used the method before 1998 [1]. Although this method has a high degree of precision, the increasing number of web pages entails using a larger number of experts, making it very difficult and impractical.

2.2 Cluster Methods

Clustering pages is used for automatic categorization of web pages. Each document is a web page and each cluster includes many documents. The first phase of clustering is extracting features. For each document the general words are first omitted. These are the words without independent separate meaning, like prepositions. Each feature is a keyword or phrase appearing in a group of documents. Keywords of a document can be extracted using different approaches like that of term frequency (tf) [9]. Next, each document is shown using a Feature Vector, which includes the features of the document and the numerical value of each feature. The numerical value shows the frequency of the feature in the document. After the formation of the feature vector, the clustering algorithm is used on the collection of vectors to categorize the documents. Examples of clustering algorithms include: BIRCH [10], CURE [11] and DCTree [9].

2.3 Structural Categorization of Web Pages

The structure of web pages is used widely to improve the organization, search and analysis of information on the web. As an example a link is intended presumably to show the topical interrelations between two documents. On this basis, the texts of web pages are divided into three groups:

- The anchor text used for description of the link
- The text near the link which usually includes around 25 words before and after the link [12]
- The regular text that forms the remainder of the page

In most web pages the anchor text and the text around the link are better descriptors of the topic of the page. Common methods of web page categorization which use word or phrases of the destination page pay attention to keywords of the page as well as the anchor text and its surrounding text to extract features. For example, Google includes the pages that have the searched keywords in texts around their links to improve search results even if the rest of the page does not have the words.

2.4 Automatic Categorization of News Using S-V-M Categorizer

The architecture of this approach includes six main modules: pre-processing, presentation, storing, S-V-M, user registration, and retrieval of web pages.

A Web page retrieval module downloads online articles and news using crawling robots from news sites. The pre-processing module includes text separators, pre-processor of documents, and generator of text vectors. The pre-processor of document omits general words and gives the remainder to the generator of the vector to make the vectors using $tf*idf$ [12]. The final product goes to the S-V-M module. Each text vector includes the remaining words of the text. There are three news databases, Reuter's tests, and the system in the storing module. The news database stores the features of the news such as date address and text of the news, downloaded from news websites. The system database stores information about users and groups related to each person. The Reuter's test battery is used to train the S-V-M web page categorizer. The next module is S-V-M.

This module is a binary categorizer including an S-V-M trainer and S-V-M categorizer to train S-V-M; a category (like sport) is selected and its related model file is produced. The model file is delivered to the S-V-M categorizer which performs categorization of the downloaded documents. The representation model shows the categorization results based on the hierarchy or priorities specified by the categorizer. The registration module manages users' information, and their personal groups.

The categorizer software performs categorization using two methods, general and specific, which are described below.

1. In the general categorization, all training documents are chosen from Reuter's documents. Ten general groups are now supported by the categorizer. An S-V-M categorizer has been produced for each group. After training, the output of the S-V-M categorizer is saved in the system database.

2. The news articles are downloaded from their sources and their extracted texts are stored in the news database.
3. When the user requests news of a special group, the recently downloaded documents are retrieved from the news database and the vectors of each document are created by the pre-processor module.
4. An S-V-M categorizer is created for each group and their model files are stored in the system database.
5. When a user requests a piece of news in his personal group, the recently downloaded pieces of news are retrieved and their document vectors are created.
6. The document vectors and the group model file are given to the S-V-M of the group and the results of the categorization are shown in order of priority.

The system is designed in such a way that when a user reads a piece of news from a specific group, he can click on the “related” button if it is the related news he wants. This feedback will be used for further training to improve the categorization [12].

3 Automatic Categorization of Persian News

In this part, the implementation of the project for categorization of Persian news is explained. This is done based on the features of Persian syntax and includes two phases. In the first phase, the features are extracted and stored in the local database. In the second phase, the similar pieces of news are extracted using the specified features. In Figure 1, the structure of the automatic categorizer of Persian News is shown.

3.1 Extraction of Features

Persian sentences include general and keywords based on their semantic load. Keywords carry the general meaning of the sentence and general words are used with keywords to complement the meaning of the sentence. In this approach, we specify keywords, the topic of the news, and its date as the features of the Persian sentences. Therefore, those pieces of news which have similar keywords and topics are regarded as related and similar.

3.2 Omitting General Words

General words are divided into two groups based on how they can be recognized in the sentence:

1. Words which are recognizable using the sentence structure.
2. Words which are specified using a list of general words

In order to find the general words based on the sentence structure, Persian sentence grammar was used as shown in Figure 2.

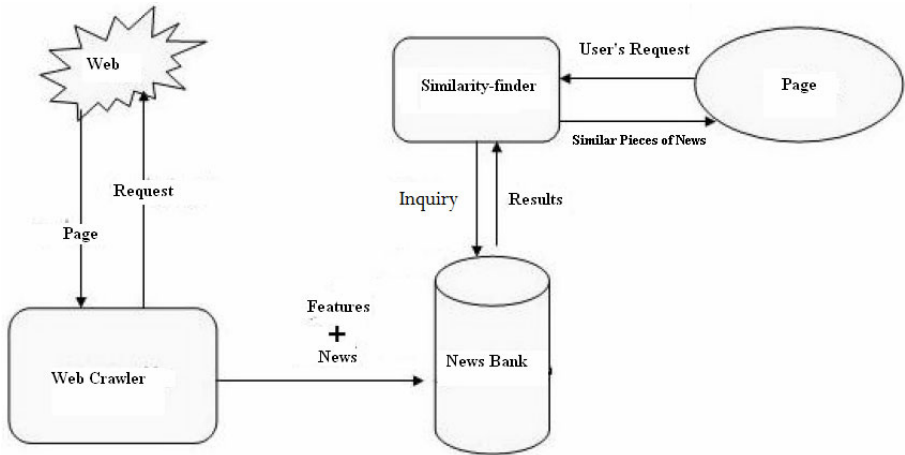


Fig. 1. Architecture of Persian news categorizer

```

<title> → <sentence> (<dot>)
<sentence>→ <word> { <space> <word> <space>}
<word>→ <general word> |<key word>
<key word>→ <letter><letter><ketter>{<letter>}
<general word>→ <verb>|<mark> |<additional word>
<additional word>→ <two word>|<other additional word>
<two word>→ <letter> <letter>
<other additional word >→ "های"|"چرا"|"این"|"زیرا"|"برای"|"همه" |...
<verb>→(<past mark>|<present mark>|<future mark>)<normal verb>
<normal verb>→<keyword>
<past mark>→ "بودم"|"بودی"|"بودند"|"بودیم"|"بودند"|"شده"
<present mark>→ "نمی"|"می"
<dot>→ "."
<space>→ " "
<letter>→ "ی"|"..."|"الف"
<mark> → ":"|"?"|";"|"",
    
```

Fig. 2. Persian sentence grammar for extracting features

The following rules can be deduced based on this grammar:

1. The word before the period in each sentence
2. The word after a conjugation of the future tense of some verbs using the auxiliary verb “خواستن”:

خواهم رفت، خواه دید

The word coming before the auxiliary verb “بودن” in the past tense, like: رفته بودم، دیده بودی

The words which are created from at most two letters are general words and are not worth categorizing and thus are omitted.

Other groups of general words include those that are used in most sentences or are not semantically worthy, so they are omitted from the news.

Omitting general words can be done in two ways:

1. Offline: after the crawler gets the news from the related site
2. Online: when the crawler requests similar pieces of news

In the first method when the crawler gets the news from the site, its general words are removed and the remaining words that are features of the news are stored in the database. In the second method, just the news is stored in the news bank and upon the user's request its general words are omitted. Then, based on the keywords, similar pieces of news are specified in the news bank. The first approach needs more memory because features of all pieces of news must be stored. But the similar pieces of news are found with higher speed. In the second method, lower memory is needed but the speed of finding similar news is reduced. As the higher speed is more significant than the memory needed, in this project, the first method was used for implementation.

3.3 Finding Similar Pieces of News

There are different definitions for similarity between the pieces of news. The similarity can be semantic or syntactic. Similarity of news is a relative concept and can be different from different points of view. For example, for some people, those pieces of news which have the same intention or content are similar, while others may consider a common word as the reason for similarity of different pieces of news.

For the project to be comprehensive, those pieces of news which have at least one similar word are considered similar. As stated before, the features of each piece of news are stored with the news in the news table. Therefore, by creating inquiries which look for pieces of news similar to the one at hand, and performing them on the news table, similar pieces of news are gathered. The inquiries are performed in order of priority on the news table and the results are shown to the user.

The priorities of inquiries for a piece of news with n features include:

The first priority: inquiries with n features

The second priority: all inquiries with permutations of $n-1$ features

The third priority: all inquiries with permutations of $n-2$ features

...

The n th priority: all inquiries including only one of the keyword.

4 Assessment

The similarity-finder software has been implemented in two parts, crawler and similarity-finder. The crawler refers to sources of news in the internet at predefined points of time and it downloads different pieces of news based on their topics and dates. In this part, the features of each news items with regard to its summary (here, the first paragraph of the news) topic, and the download time are stored in the site's local database.

In the second part, the similarity-finder, the news items similar to the particular one are searched and retrieved. This part, first, extracts the features of the selected piece of news, based on which it creates appropriate inquiries and looks for similar news in the news database. These inquiries are created based on the priorities and permutations of features of the news.

The crawler system and similarity-finder are implemented in Linux operating system using PHP language, some of whose results are represented in Figure 3 below.

<u>وزیر صنایع و معادن ایران عازم کشور سوریه شد</u>	اخبار مشابه برای خبر :
<p><u>وزیر صنایع و معادن ایران عازم کشور سوریه شد</u> - خبرگزاری جمهوری اسلامی ایران - ۲۰ ساعت و ۵۰ دقیقه قبل</p> <p><u>وزیر صنایع و معادن ایران عازم کشور سوریه شد</u> - خبرگزاری دانشجویان ایران - ۲۲ ساعت و ۲۵ دقیقه قبل</p> <p><u>معاون وزیر صنایع و معادن به عنوان مدیرعامل جدید شرکت آلومینیوم ایران معرفی شد</u> - خبرگزاری دانشجویان ایران - ۲ روز و ۶ ساعت قبل</p> <p><u>وزیر صنایع و معادن خواستار شد: انتقال دانش طراحی بدنه خودرو از سوی رنو به سایپا و ایران خودرو</u> - خبرگزاری دانشجویان ایران - ۳ ماه و ۷ روز قبل</p> <p><u>وزیر صنایع و معادن به نروژ رفت</u> - خبرگزاری دانشجویان ایران - ۳ ماه و ۷ روز قبل</p>	

Similar pieces of news for: The Iranian Minister of industries and mines left for Syria

<p><u>The Iranian Minister of industries and mines left for Syria</u>-Islamic Republic News Agency - 20 hours and 50 minutes ago</p> <p><u>The Iranian Minister of industries and mines left for Syria</u> - Iranian Students News Agency- 22 hours and 25 minutes ago</p> <p><u>Deputy Minister of industries and mines was introduced as the new manager of Iranian Aluminum Company</u> - Iranian Students News - 2 days and 6 hours ago</p> <p><u>Minister of industries and mines suggested: transfer of knowledge of designing car body from Renault to SAIPA and Iran Khodro</u>- Iranian Students News - 3 months and 7 days ago</p>
--

Fig. 3. Persian sentence grammar for extracting features

Evaluating the results obtained by the software and their manual examination showed that using the permutations in headline, summary and topic of news has a precision of 79 percent. This figure is the result of manual examination of 100 pieces of news and similar news items and comparing them with the results obtained by the similarity-finder software for these pieces of news

5 Summary and Future Work

In this article the implementation of similarity-finder software for Persian news is elaborated on. First, different methods for categorization of web pages and web news are examined. Then the method implemented for Persian news is elaborated on. Finally, the results of implementing the software and its manual examination are presented.

Since in this system only the resemblance in subject and keywords in the news have been used, the results are not 100 percent accurate. A method of solving this problem to obtain more accuracy in the search results is to use semantic similarities to find similar pieces of news. For this to happen, the features must be chosen for both keywords and concepts of the news. The semantic features may have little in common with keywords and just take into account the meaning of a piece of news. For example, if there is the word "برانکو" (the former coach of Iran's football team) in the news text, the pieces of news about Iran's football team must be extracted as well. But, based on keywords, only those pieces of news which have similar words and general topics are presented as similar without regard to semantic similarities.

With regard to the results of evaluating the similarity-finder software based on keywords, it is expected that using semantic similarities helps us obtain a high precision - above 79%- in finding similar pieces of news and categorizing them. Implementation of a semantic similarity-finder can be the next generation of categorizers of Persian news as compared with the similarity finders based on keywords.

References

1. Kwon, O.-W., Lee, J.-H.: Web Page Classification Based on K-Nearest Neighbor Approach. In: Proceedings of the Fifth International Workshop on Information Retrieval With Asian Languages, pp. 9–15. ACM, New York (2000)
2. Yang, Y., Lui, X.: A Reexamination of Text Categorization Methods. In: Proceedings of the 22nd International Conference on Research and Development in Information Retrieval (SIGIR 1999), pp. 42–49. University of California, Berkeley (1999)
3. Lewis, D.D., Ringuette, M.: A Classification of Two Learning Algorithms for Text Categorization. In: Third Annual Symposium on Document Analysis and Information Retrieval (SDAIR 1994), Las Vegas, USA, pp. 81–93 (1994)
4. McCallum, A., Nigam, K.: A Comparison of Event Models for Naïve Bayes Text Classification. In: AAAI 1998 Workshop on Learning For Text Categorization (1998)
5. Combarro, E.F., et al.: Introducing a Family of Linear Measures for Feature Selection in Text Categorization. *IEEE Transactions on Knowledge and Data Engineering* 17, 1223–1232 (2005)
6. Apte, C., Damerau, F., Weiss, S.M.: Automated Learning of Decision Rules for Text Categorization. *ACM Transactions on Information Systems* 12, 233–251 (1994)

7. Dumais, S.T., Platt, J., Heckerman, D., Sahami, M.: Inductive Learning Algorithms and Representations for Text Categorization. In: Proceedings of the Seventh International Conference on Information and Knowledge Management (CIKM 1998), pp. 148–155. ACM, New York (1998)
8. Weigend, A.S., Weiner, E.D., Peterson, J.O.: Exploiting Hierarchy on Text Categorization. *Information Retrieval* 1, 193–216 (1999)
9. Guha, S., Rastogi, R., Shim, K.: Cure: An Efficient Clustering Algorithm for Large Database. In: Proceedings of the ACM SIGMOD Conference on Management of Data, pp. 73–84. ACM, Seattle (1998)
10. Tokunaga, T., Makoto, I.: Text Categorization Based on Weighted Inverse Document Frequency, Special Interest Groups and Information Process Society of Japan, SIG-IPJS (1994)
11. Zhang, T., Ramakrishnan, R., Livny, M.: BIRCH: An Efficient Data Clustering Method for Very Large Databases. In: Proceedings of the ACM SIGMOD Conference on Management of Data, Montreal, Canada, pp. 104–114 (1996)
12. Chan, C.-H., Sun, A., Lim, E.-P.: Automated Online News Classification With Personalization. In: Center for Advanced Information Systems. Nanyang Technological University Nanyang Avenue, Singapore (2001)

An Evaluation of TS13298 in the Scope of MoReq2

Gülten Alır¹, Thomas Sødning², and İrem Soydal³

¹ Aksaray University, Faculty of Education, Aksaray, Turkey
gulten.alir@aksaray.edu.tr

² Oslo University College, Faculty of Journalism, Library and Information Science,
Oslo, Norway
thomas.sodning@jbi.hio.no

³ Hacettepe University, Department of Information Management, Ankara, Turkey
soydal@hacettepe.edu.tr

Abstract. TS13298 is the first Turkish standard developed for electronic records management. It was published in 2007 and is particularly important when developing e-government services. MoReq2, which was published in 2008 as an initiative of the European Union countries, is an international “de facto” standard within the field of electronic records management. This paper compares and evaluates the content and presentation of the TS13298 and MoReq2 standards, and similarities and differences between the two standards are described. Moreover, the question of how MoReq2 can be used as a reference when updating TS13298 is also dealt with. The method of hermeneutics is used for the evaluation, and the texts of TS13298 and MoReq2 were compared and reviewed. These texts were evaluated in terms of terminology, access control and security, retention and disposition, capture and declaring, search, retrieval, presentation and metadata scheme. We discovered that TS13298 and MoReq2 have some “requirements” in common. However, the MoReq2 requirements, particularly in terms of control and security, retention and disposition, capture and declaration, search and presentation, are both vaster and more detailed than those of TS13298. As a conclusion it is emphasized that it would be convenient to update TS13298 by considering these requirements. Moreover, it would be useful to update and improve TS13298 by evaluating MoReq2 in terms of terminology and metadata scheme.

Keywords: TS13298, MoReq2, electronic records management standards.

1 Introduction

The phenomenal increase in the amount of electronic information that is being generated has made the problem of information management a current issue. In a study carried out by the Association for Information and Image Management [1], some notable results regarding the issue of information were presented. According to the study, while there was an increase in the number of print records in 56% of the organizations surveyed, there was a decrease in the number of such records in 22% of the organizations. The report found that the number of electronic records had also increased in 70% of the organizations, without a corresponding decrease in the number of such records in any of the organizations. This shows a trend that the overall

amount (print and electronic) of records being generated is increasing while the amount that are print based is decreasing and the amount of electronic records is increasing. According to the same study, 37% of organizations are scanning over half of their incoming (print) documents and 12% of organizations scan more than 80% of incoming (print) documents. Many of the organizations (60%) that do not use an electronic records management system (ERMS) are experiencing problems with electronic records management. These organizations do not feel confident about the security of their records in terms of change, deletion or inappropriate access [1].

The applications used in the public sector (in particular) must support the policies governing freedom of information, while efficient information sharing is dependent on the use and evaluation of reliable information. In addition to this, success in basic essential operations must be achieved such as ensuring the reliability and legal validity of records, as well as data protection. However this is dependent on supporting the organization with an effective electronic records management. A sound structured operation of public management can only be realized with the support of records management. A sound records management system is the foundation for transparency within the state, and also an important factor for the state to be able to perform its responsibilities [2], [3], [4].

Records management operations have developed in line with technological innovations and have only recently been evaluated within the archive discipline. Efforts made by national and international societies in recent years have gained a new dimension with the procedures, standards, and applications that have been developed. Electronic records management standards have been established based on research by both national and international organizations. As a result of these standards, there is a general expectation of increased efficiency by public organizations in their e-government operations. This will be achieved by providing uniformity of the data structures, providing information/data interchange at national and international level, and by developing documentation that will instruct organizations generating electronic records and software developers within the field of records management.

It is inevitable that standards need to be reviewed, updated, and harmonized with the new conditions and situations over time. Standards do have a life cycle as well and can never be considered to be complete [5]. Waalwijk [6] also underlines that standards only have a limited lifetime and need to be revised.

Turkey also needs to maintain its efforts and work in the field of electronic records management which started with e-government applications. It is also crucial to update the standards, the most significant part of these efforts, in line with international and technological developments. Employed for the establishment of TS13298¹, MoReq has a significant place with its updated new version. Particularly, such a standard supported by the European Commission is also significant for Turkey aiming at adopting the European Union norms in the long term to act in line with developments in Europe.

MoReq2 (Model Requirements for Management of Electronic Records) is the result of a study to which numerous experts from different disciplines contributed. Moreover, it has been taken as a basis by several countries for developing their national standards. MoReq2 has not been prepared for any commercial purpose or

¹ In this study the term “TS” stands for “TS13298”.

according to the needs of any particular country or organization, which increases its applicability for international development and acceptability of the standard for national applications.

The purpose of this study is to evaluate the TS standard, which plays a key role in electronic records management in Turkey, within the framework of the MoReq2 standard. Moreover, the aim of the study is also to give a clear overview of the similarities and differences of the two standards, and to present recommendations for additions, arrangements, and updates which can be introduced into TS. The scope of MoReq2, which is a model document, is much wider than TS. On this point, it is considered whether MoReq2 with more developed specifications can make contributions to the development of TS or not. Taking specifications and content of an internationally accepted standard as a reference will present a definite advantage in terms of time and effort. MoReq [7] was used when developing TS. Therefore, it is of great importance to evaluate the developments in MoReq2 against TS. The aim of this study is to increase awareness concerning the developments in MoReq2.

2 Standards and Records Management Standards

A standard is defined as a “*document, established by consensus and approved by a recognized body, that provides, for common and repeated use, rules, guidelines or characteristics for activities or their results, aimed at the achievement of the optimum degree of order in a given context*” [8]. There are two types of standards [9]:

- De jure standards: those officially issued by governments or standards bodies.
- De facto standards: those not issued by such bodies but nevertheless considered valid standards because they have been widely adopted by a community of users.

Standards, technical reports, guidelines, and white papers create a professional environment of "best practice" procedures. The purpose of standards is to ensure conformity among products in all countries where companies developing those products do business. The first priority with regards to users is to ensure high quality and performance in goods and services. Products developed against standards ensure interoperability between the competing organizations, and objective criteria, on which agreement is reached, can be used when defining and testing quality. At the same time, safety, consistency, uniformity, reliability and economic efficiency can be ensured with standards. Standards and guidelines are fundamental reference tools not only for the development of specific tools, but also for ensuring and maintaining quality in non-concrete fields such as the content of information and workflow. Standards also allow companies to agree on common values that can cross local, national and even international lines [10], [11]. Like the other disciplines of the information management field, records management is affected by developments in information technologies. New standards are established, the existing ones are developed and management and system approaches consistent with standards are adopted. Public and voluntary organizations have started to develop standards, and to use physical and managerial systems, for various reasons, such as requirements as a result of the development of e-government applications

(freedom of information act, data protection act, etc.), or the fact that there is a need to increase the efficiency and quality of information management while decreasing its costs.

Hofman [12] states that records management standards will raise the level of professionalism as well as the level of authority, and will provide organizations as well as the records management profession with a comprehensive framework.

Public organizations such as the International Standards Organization (ISO), American National Standards Institution (ANSI) and voluntary organizations, such as ARMA International, International Council of Archives (ICA), Document Lifecycle Management Forum (DLM Forum) have carried out international studies on establishing records management standards. While some of these standards are mainly about specific issues, namely preservation, e-mail management, metadata², some of them cover records management applications and general procedures. For instance, “ISO 15489 Information and documentation -- Records management” [13] is the first international standard where general procedures in records management are underlined [14]. It has provided the very basis for various national and international standards in the field of records management for determining common requirements of terminology, characteristics of records, records management process, and workflow.

MoReq [7] is the first international standard covering functional requirements/specifications expected to be available in specialized software required for establishing Electronic Records Management Systems (ERMS). MoReq has guided various national standards and has been translated into several languages.

2.1 Model Requirements for the Management of Electronic Records (MoReq2)

MoReq2 was created by DLM Forum (Document Lifecycle Management Forum) through an initiative of European Commission countries. Established in 1997, DLM Forum has been an independent organization since 2002. The main purposes of the DLM Forum include “*to establish a strategic network to bring together individuals and organizations, from both the public and private sector involved in Document Lifecycle Management; ... to raise awareness in the field of Document Lifecycle Management; to provide technology and knowledge transfer and information services, appropriate practice guidelines, benchmark indicators and information, educational, skills development and research opportunities*” [15]. Although founded as a European-centric organization, DLM Forum is available for all organizations and researchers from all around the world.

MoReq was developed by DLM Forum and published in 2001. It was composed of 390 requirements and 100 pages [7]. Due to technological advancements, this document needed to be improved and updated. As a result of the research that was undertaken, a new version of MoReq called MoReq2 was published in 2008 [16]. The development in MoReq2 is qualified as an “evolution” in the literature. MoReq2 provides an approach for the implementation of records management principles independent from any format or system. Some of the around 800 requirements in MoReq2

² For detailed information, <http://www.ica.org/en/standards>,
<http://www.arma.org/standards/index.cfm>, <http://www.iso.org>

are mandatory, while some of them are only optional.³ MoReq2 has been established as a superset, which can meet requirements of a big organization at maximum level. Thanks to its modular structure, organizations with small structures can also make use of the requirements of the sections needed as well. Some sections of MoReq2 (Sections 3 to 9) contain the core ERMS functional requirements in detail. Section 10 is defined as an “Optional Module”. Some of these sections will be essential for some organizations, but unnecessary for others. In this way, added value is provided by taking into consideration the needs of organizations with different dimensions, “Section 11” contains non-functional requirements and “Section 12” identifies requirements for managing metadata [16].

An ERMS may be certified MoReq2 compliant by an accredited MoReq2 test centre. Among the characteristics of MoReq2 there are testability, metadata scheme and XML scheme. MoReq2 was designed and written with testability, and testing materials have been developed and published alongside the electronic copies of the requirements. The Metadata scheme of MoReq2 is much more detailed than MoReq in terms of both scope and presentation. Each data element is expressed in tables in MoReq2 where 158 unique metadata elements are defined. The name of the metadata element, its obligation status and definition, whether more than one value is allowed for the element (occurs), how the value(s) for this element are produced (populated), the suggested default value (default), rules for the inheritance of the metadata values (inheritance), conditions and rules that govern the use and value(s) of the element (use conditions), references to formal requirements (requirements) from MoReq2 that can change values of the metadata element, etc are described in detail [17]. The XML scheme developed in MoReq2 provides a way for electronic records, with their metadata, to be transferred between systems without loss of their functionality. A significant advantage of MoReq2 is that countries can make translations and amendments in accordance with their own languages, terminologies, and laws in the section called “Chapter 0”. MoReq [7] has been a model for various national standards, and 10 countries have published this standard in their own languages. Moreover, it is one of the reference documents used for the development of TS and NOARK5 [18]. MoReq2 had also been translated into four languages (French, Russian, Czech and Slovenian) as of January 2010 [19].

In spite of these developments, criticisms of MoReq2 focus on the fact that it is highly detailed and long and that it may be problematic for small companies to cover the cost of meeting these requirements [20], [21]. Indeed, the reason why this document is so long and detailed is that it has been prepared foreseeing at the highest level the possible challenges of a large system. However, it also covers specifications to meet the requirements of small-sized organizations. MoReq2 continues to be developed taking into consideration the various criticisms against it. Among the 2010 objectives stated by DLM Forum December 2009, is the establishment of a core module composed of mandatory requirements [22].

³ “MoReq2 contains both mandatory and desirable requirements. This level of mandation is indicated as follows:

- the word “must” indicates that a requirement is mandatory;
- the word “should” indicates that a requirement is desirable.”

In all cases, the level of mandation is dependent on its context” [16].

MoReq2 is an independent document prepared with the contributions of several experts from various countries and disciplines, supported by the European Commission, dealing with the requirements of both public and also private organizations in detail. This and other similar characteristics of the standard bring out the question of “national standards or MoReq2” for several countries of the European Union and other countries. As stated in MoReq2 as well, countries have their own peculiar laws, characteristics, records management traditions and needs. However, these specifications do not pose any obstacle to evaluating or adapting the document presented by MoReq2. As a matter of fact, as stated above, certain countries have already translated MoReq2 into their own languages. Moreover, there are various studies in which MoReq is compared to national standards in the literature. In these studies, generally a method based on evaluating the texts is employed, and in some studies there are interviews with the persons dealing with the issue.

Wilhelm [21] has evaluated the differences between TNA 2002 of the United Kingdom and MoReq2. At the same time, he has carried out interviews with at least one representative of each of the key stakeholders in order to discover their opinions. According to the study, MoReq2 has similarities with TNA 2002 in terms of content, structure and expression, but MoReq2 provides clear and unambiguous definitions particularly of ERMS and optional modules, and provides a more up to date, comprehensive content. However, there are criticisms that the objective of the standard is too large, and that therefore the length of the document is problematic, and that it covers too numerous requirements.

Wilhelm [23] has also evaluated the interviews for the comparison of MoReq2 and TNA 2002. He pointed out that the most significant problem concerning the implementation of MoReq2 is that it is a highly detailed and long text.

A similar study was carried out by Henttonen [24]. Henttonen compares MoReq2 and Finnish SÄHKE specifications to see their differences and similarities. The study, carried out by comparing the texts of the two standards, seeks to find out whether there is an easy way to harmonize them. According to the conclusion of the study, MoReq and SÄHKE have considerable differences in terms of data models, elements and functional requirements. In spite of some similarities, MoReq2 is more technical and detailed than SÄHKE, and the two standards are quite incompatible. However, as a conclusion, Chapter 0 of MoReq2 is referred to and it is emphasized that MoReq2 can meet the needs at national level if SÄHKE is abandoned.

On the other hand, while Waalwijk [6] gives an account of the establishment process of the requirements for electronic records management within the Netherlands Institute for Standardization, he underlines the place of MoReq in this process and makes comparisons between MoReq, ReMANO and the Core model. The study pointed out that ReMANO and the Core model were taken as the basis in developing Netherlands standard NEN 2082, and that MoReq needed to be developed and reviewed as soon as possible and to be published once again as MoReq2.

2.2 TS13298 Information and Documentation-Electronic Records Management

The idea of preparing a document which will guide the organizations for the electronic records management in Turkey emerged in line with e-government work. The study mentioned in Article 37 of the “e-Transformation Turkey Project 2005 Action Plan”, was presented to the public for comment for the first time in 2005 as “Reference Model of Electronic Records Management System Requirements (v.1.0)”.

The document is arranged as a standard inline with the recommendations and issued as a Turkish standard in 2007 [25]. The standard is an important source for both organizations and software companies as it defines the records management requirements, and includes instructions regarding the freedom of information act and e-signature law. It also underlines the importance of building awareness in the field of records management and of pursuing the policy of records management within all units of an organization. In compliance with the regulation dated 16 July 2008, governmental institutions are required to adapt their electronic records management system in accordance with TS [26]. The text of the standard document, produced for defining functional specifications required to be available in ERMS software, is composed of 42 pages and three sections: system requirements, record requirements and metadata scheme. The main subjects such as classification, retention schedule, registration operations of electronic records, specifications for use, access control and security, and system design, are covered in the first section. In the records criteria section, there are records specifications, document management, non-electronic systems, and digital scanning systems. The last section covers metadata elements. It is composed of 93 metadata and requirements listed in 11 articles. There is no information concerning whether metadata is mandatory or optional. However, the requirement on general specifications (TS11.1.d)⁴ states that “The system manager will be able to determine whether metadata elements are mandatory or optional in an ERMS system and whether these elements will be indexed or not during the system design stage”.

3 Methodology

The hermeneutics method was employed in comparing TS and MoReq2. According to Walker, the hermeneutics study “...attempts to make meaning of words and narrative and to gain understanding, it is a powerful tool for studying intact text. Methodologically, the different sources of data were analysed using a selective reading approach” [cited in 27]. Cole and Avison [28] defined the hermeneutic method process as

A piece of text is read several times asking: what statement(s) or phrase(s) seem particularly essential or revealing about the subject’s prejudices and/or the phenomenon or experience being described?...Concepts help to maintain the focus of analysis on the circular presupposition of hermeneutics (that the whole precedes the determinant of the details and through

⁴ The section and subsection names and the requirements will be mentioned either with the initials of TS (for TS13298) or with the initial of M (for MoReq2) further on in the text, in order to be able to distinguish these standards more easily.

the details we construct the whole). They also help to integrate different data sources by treating them as a singular totality. [28]

There are studies in the literature where ERMSs are compared employing hermeneutics method [21], [23], [24].

In the light of this method, comparing and evaluating MoReq2 and TS is based on the texts of these two standards. When evaluating the standards, primarily their general specifications were taken into consideration, read, and the sections to be compared defined. These sections, the subsections under each section, and requirements listed under subsections were read many times and compared. As it is not possible to specify all sections and requirements individually within the scope of this study, certain sections and requirements of particular importance for the subject, were dealt with.

TS and MoReq2 were evaluated comparatively with regard to the requirements under the sections of *content*, *terminology*, *access control* and *security*, *retention* and *disposition*, *capture* and *declaring*, *searching*, *retrieval* and *presentation* and *meta-data scheme* (see Appendix). The sections were defined taking into consideration the purposes of the documents with regard to general scope, form of expression of the requirements, approach to the subject, and presentation specifications, and also samples were given with respective requirements. The study also addresses the question “Are there any specifications of TS which need to be improved and/or updated compared to MoReq2?”.

4 Findings and Discussion

The TS is composed of 42 pages and three sections. It is stated in the document that the standard defined “the minimum functional requirements which need to be present in ERMS software”. As there is no expression to the contrary, it will be useful to evaluate all requirements in the document as mandatory. Moreover, definitions regarding the records specifications are given in the second subsection of the standard (TS7.1). However, only six terms were defined for the terminology section of the standard and for the definitions of other terms, references were made to ISO 5127 and 15489. This structure of the document leads to ambiguity.

MoReq2 is composed of 207 pages and eleven sections. The first two sections cover information and terminology with regard to the standard. Moreover, it is supported with an XML scheme. In the second section of MoReq2, there are detailed definitions under the “2.1.Key Terminology and 2.2.Key Concepts” subsections, and there is an extensive glossary (13. Reference Model) in the last section.

Controls and security: TS gives the definitions of user roles and groups in access control and security, and presents requirements of basic definitions suitable for these roles. In both standards, emphasis is laid on having access restrictions suitable for the user roles and the authority of administrator/system administrator. As for TS, with the “TS5.2.a, TS5.2.b, TS5.3.b, TS5.4.a, TS5.5.a, TS5.7.f” requirements authorization is given to the system administrator. In MoReq2, the explanations of user roles and access

are given in subsection “4.1.Access” and mainly define in detail the authorizations of the role defined as “administrator”. In this subsection, 14 of the 24 requirements are within this scope. Considering TS as a national standard, the access rights for the records and user roles (system administrator, records manager, etc.) are defined at a minimum level. This supports the purpose of the document regarded as a guide for organizations, which are to meet the minimum requirements. However, defining the authorizations of the administrator in a certain manner with regard to such a subject is important. In these types of standards, there have to be requirements leaving no ambiguity with regard to role definitions and responsibilities. The table available in MoReq2, which explains the user roles, [16] proves very useful for defining authorizations.

Audit trails evaluated within the control and security section allow for recording of the user and administrator roles in the system and their activities. “Audit” is available in subsection “TS5.9” in TS. TS employs simpler and clearer expressions than MoReq2. Requirements for changing MoReq2 audit trail configurations (M4.2.4) and recording these parameter changes (M4.2.5) are defined as well. However, MoReq2 also provides significant requirements for exporting data on any element or user stored in audit trail in a safe manner (M4.2.15). In MoReq2, requirements concerning backup and recovery, which will allow recovery of records if any are lost because of system failure or accident (M4.3), and vital records, which are considered absolutely essential to the organizations, (M4.4) are stated under various sections. As for TS, the issue of storing the data is addressed (TS6.4), but no requirement concerning backup is given. Only under the document management section with the subsection “8.5 Storage and Backup”, is there a single requirement which states “Electronic documents must be copied in safe storage and back-up units”.

Retention and disposition: One of the most important elements of a records management system is retention and disposition. These operations, which need to be performed in accordance with a policy, must be addressed fastidiously. The retention schedule of TS (TS2) is composed of six subsections. In the general rules, there are 11 requirements in total concerning defining a retention schedule during the system installation, reporting and disposition function, role of records manager, association with class, elements of retention schedule, automatic notification. The storage times subsection (TS2.4) is composed of two requirements, which offer some alternatives for the beginning dates, as well as the minimum and maximum durations for the storage time of the records. Moreover, under this subject there are also certain definitions such as hierarchy of classification (TS2.2), criteria of retention and disposition process (TS2.5).

MoReq2 defines the retention and disposition subject in a detailed manner avoiding any confusion. The administrative role for changing or deleting a retention and disposition schedule (audits and metadata) is clearly defined (M5.1.8, M5.1.23). Furthermore, the requirements concerning data transfer are listed under subsection “5.3.Transfer, Export and Destruction” in MoReq2. As for TS, the requirements concerning collective transfer operations are stated without any detail with two requirements listed under electronic records registration operations (TS3.1.s, TS3.1.t).

Capturing and declaring records: The process of capturing the records means not only inserting the records into the system, but also making necessary definitions for

the short and long term preservation of the records. Registration of records covers important stages such as the mapping of records to its classification scheme, definition of access rights, etc. Importance is attached to this issue in both standards (TS3. Capturing of electronic records, M6.Capturing and declaring records), but details of MoReq2 once again come to the fore. Both standards underline the fact that the system must support all types of records. Ten of the requirements of MoReq2 listed under the section of “Capturing and Declaring”, are common to some requirements in TS3, such as capturing of electronic records (M6.1.1/TS3.1.c), components (M6.1.3/TS3.1.d), metadata requirement (M6.1.11/TS3.1.d). While MoReq2 gives each condition as a separate requirement, more than one requirement is joined under a single statement in TS. This also applies to the specifications listed under requirements. While MoReq2 avoids revealing the names of the file formats which can be included in the system, it gives a general list of record types which are frequently used (M6.1.1). These lists may be useful for the system designers to know more about the existing record varieties. There is no such general list in TS. Requirements concerning metadata, audit trail (M6.1.6, M6.1.9, M6.1.14, M6.1.17, M6.1.18, M6.1.21, M6.1.22) and keyword definition (M6.1.23-28) are defined in detail in “6.1: Capture” subsection of MoReq2.

E-mail management: E-mail is another significant type of document for organizations. There are certain problems both in transferring them to ERMS and in defining e-mail messages, which have different components than known record specifications. Both standards present alternatives for transferring e-mails (TS3.4.a/ M6.3.3). While MoReq2 specifies 18 requirements for definitions of metadata and the process of transferring them to the system in addition to these alternatives, TS simply states that e-mails must have a format, in which owners of e-mail addresses can be easily understood.

Scanning and Imaging: The operation of digitizing records which are not in electronic format, and transferring them to ERMS is given in the subsection “6.5.Scanning and Imaging” in MoReq2. In TS, these operations are given in the second section “Record Criteria” with the section “10. Digital Scanning Systems”. TS mainly focuses on the digitization specifications of the records transferred to digital form which are originally produced in print form. While both standards declare common requirements for not limiting the system to certain formats, TS also specifies some resolutions that are suitable for different types of materials, coloring options, common file formats and their specifications (TS10.4, TS10.5, TS10.9). Similarly, the copying of records (master copy, usage copy, mini copy) is given in subsection “10.3. Scanning Techniques”. There has been a growth in the number of organizations that digitize their archives in Turkey in recent years. This information can be explanatory and guiding for organizations and system designers. However, rapidly developing technology brings new formats and new techniques. Therefore, this information must be updated from time to time. For instance, PDF-A format stated in MoReq2 (M6.5.3) has become widespread in recent years. According to the AIIM’s (2009) study, 56% of scanned documents have been switched from the traditional TIFF format, with a 12% adoption of PDF-A, which is a more suitable archive format.

Search, retrieval and presentation: When evaluating the two standards in terms of search, retrieval and presentation, they both have common requirements in basic issues such as use of full text and/or metadata, generating combinations, controlled vocabulary, advanced search and retrieval of search results without any further operation (TS 4.2/M8.1). As seen under the subsection “8.1: Search and Retrieval”, when defining a requirement in MoReq2, reference is made to the respective standards, if any. For instance, in requirements M8.1.17 and M8.1.18, when a thesaurus is included in ERMS the standards which must be followed (ISO 2788, ISO 5964) are also stated. As seen in requirements M8.1.21, M8.1.22 and M8.1.29 of MoReq2, advanced search functions are explained in detail with samples. Similarly, print options are extensively dealt with (M8.3.7- M8.3.16), and some handy requirements are listed. On the other hand, TS (4.4) presents three requirements for print options. Although explanations in MoReq2 seem to be more detailed, they turn out to be useful in avoiding problems which may appear in time.

Metadata scheme: Metadata scheme and XML is one of the strong features of MoReq2. Taking into account metadata requirements, several basic requirements are in common, such as limitations concerning metadata elements (M12.2.1/TS11.1.a), functions of metadata and ERMS (M12.2.2/TS11.1.b), formats which metadata elements must support (M12.2.3/TS11.1.c). However, details in the requirements of MoReq2 (M12.2.9, M12.2.16) and references to the standards (M12.2.7, M8.1.17, M8.1.18) come to the fore as well.

Metadata elements are the proof of the life-cycle of records. Therefore, they must be formed with great care. Metadata requirements of organizations can vary as well. However, a standard must be guiding and explanatory in terms of both content and presentation. MoReq2, Appendix 9 [17] presents the metadata elements in a very practical format (See Figure 1). Even this document can be an important guide for the system designers. Information concerning the location of metadata in the system, how it will be defined, under what conditions it will be used, the changes which can be made, and values which can be defined is given in a systematic and clear manner. Such a structure is important with regard to deciding and describing the metadata required.

93 metadata elements are defined in TS “Section III: Metadata Scheme”. Metadata is composed solely of metadata name, short definition and short explanations below some of them. There is no systematic structure. Given that electronic records management in Turkey does not go back very far, it is needless to say that there is a lack of information and know-how concerning metadata in both public and private organizations in Turkey. Therefore, TS is a quite significant reference source. As noted earlier, evaluating all metadata elements individually is not included within the scope of this study. However, it is considered that the metadata structure of MoReq2 in terms of presentation and scope could be an important reference model for the improvement of TS. Thus, it will be possible to present organizations with a more proper and standard structure to select their metadata elements.

	Obligation:	Mandatory	Occurs:	Once
Definition:	A name given to a record type.			
Populated:	Manually entered when the entity is created.			
Source:	User.			
Default:	None.			
Use conditions:	Can be changed by an administrative role or by users who appropriate access rights. Must be unique.			
Comment:	No comment.			
Requirements:	5.1.15			

Fig. 1. MoReq2 example of metadata scheme [17]

5 Conclusion and Recommendations

Developed in context of e-government work, TS is one of the most significant steps in electronic records management in Turkey. It has filled an important gap as a guide for organizations and companies producing software. The rapid development in technologies and the changing conditions make it necessary for standards to be improved. Established as a national standard for the first time, it is quite correct and natural for TS to present minimum requirements at the first stage. However, when the purpose is to generalize the use of standards, taking into account the diversification of requirements of the organizations, TS has to be further developed. One of the important issues to develop TS in terms of terminology is building a glossary section. Giving explanations concerning the subjects in sections, if deemed necessary, will ensure the integrity of the document and will make it easy to use. Providing definitions of the requirements concerning access control and security must not leave any ambiguity in terms of system security and management. Particularly, to define the responsibilities of the system administrator more clearly and in more detail is important for the development of TS in terms of its terminology. At the same time, the classification of the authorizations of user roles and presenting it in a table like in MoReq2, [16] will make it easy for the roles to be defined. Arranging audit trails, presenting more detailed requirements related to the records and adding more requirements concerning the storage of the data, need to be taken into consideration as well.

Widening the scope of retention and disposal requirements will pave the way for the control of this operation, which is of crucial importance for the organizations. Particularly, clarifying the responsibilities of the records manager and system administrator, and adding requirements for disposal process with metadata and audit trail, are needed. Furthermore, requirements which should be included concerning transform and export can be taken into consideration particularly to prevent data losses which may occur in the case of merging organizations.

Expanding the scope of metadata and audit trail requirements concerning registration of records will have an effect on the long term preservation of records. Furthermore, an extensive list of possible electronic record types can be established, and record types like blogs, wikis, instant messaging systems, etc. can also be taken into consideration.

E-mails are one of the fundamental tools of communication for some organizations. It is obvious that TS needs to be reviewed in terms of inclusion of e-mails in the system and providing access to the records components by defining them. The questions of what components of e-mail (attachment, message, etc.) will be defined and how, their classification, definition of their metadata, providing access for the attachments as well as the messages can be dealt with in more detail.

Operations concerning the digitization of records may be the field where other standards are also referred to the most. The operations performed in this process for long term preservation and access of the records and the formats used is of great importance. We believe that tables on file formats in TS are useful for many users. However, these kinds of formats develop constantly and new products are generated. In this context, reviewing sections on file formats will be useful for the adaptation to the new technologies.

For TS, it is possible to express the searching options in a clear manner. It will be useful to add guiding requirements for using a controlled vocabulary or thesaurus for "subject" searching, which is an important component for the records in Turkish. Furthermore, the existing requirements can also be extended in order to develop print options during the system installation.

Metadata elements are the basic management elements of an ERMS. A well-defined metadata scheme is important for the proper operation of the system. Converting the metadata in the TS metadata scheme into a more systematic and explanatory format will be useful for both users and system designers, and will contribute to the compliance of records with the standards. The status, scope, terms of conditions, values of the metadata elements and whether any alterations need to be made for these elements have to be included in TS, with their explanations. Furthermore, a systematic metadata scheme, which will be established by taking the MoReq2 metadata scheme as a sample, will be useful for the ease of use of TS. Working on the establishment of an XML scheme, which can be used for the transfer of data between systems, will be an important step for the improvement of the standard.

It can be said that the main reason why there are more differences between TS and MoReq2 than similarities is that their purposes and the groups which they address are different. However, this is not an obstacle to improving TS by taking MoReq2 as reference. MoReq2 has already been prepared as a reference document and designed to be a superstructure. As it has not been prepared considering solely a specific organization or country, it addresses a broader user audience. Although it is criticized in terms of its details and cost, it aims to build the best system possible with a content aiming to eliminate the possible problems which may occur in the system, giving explanations and references concerning the issues. Planned to be revised in 2010, MoReq2 sends correct signals to be an important standard to be adopted by many countries in the future. As a candidate country of the European Union, Turkey must take the progress of MoReq2 into consideration and revise its national standard. In this context MoReq2 offers some good options like Chapter 0. Turkey should con-

sider creating a Chapter 0 or update and revise TS by mapping MoReq2 in terms of its structure and requirements. The pros and cons of these alternatives should be discussed and examined in detail in future studies.

Indeed, standards cannot be a stand-alone solution to all problems. However, raising awareness in records management, which is a very new subject for Turkey, filling the gap in the field in terms of information and experiences, developing and generalizing the use of standards is of great importance in order to act in parallel to future technological developments. It is clear that it will be vital to improve and update TS given that all of the governmental institutions will have adopted their electronic records management system in accordance with TS [26]. This process must be undertaken not only by record managers but with the joint experience and know-how of actors in various fields, such as organization managers, users, data processing experts, computer engineers. Thus, participation from all fields will be achieved for the developments, and it will be possible to establish ERMS in Turkey on firm ground. These developments will contribute to an enhancement of e-government, e-business services.

Comparing and evaluating all metadata elements individually is not within the scope of this study but can be a guide for more specific studies. A detailed study on metadata schemes is being carried out within the scope of TS, MoReq2 and NOARK5 (Norwegian standard for ERMS) with the support of Oslo University College and the Research Council of Norway. It is planned that the study will be completed in April 2010. It is believed that this and similar studies will provide significant contributions for building awareness on ERMS and for the development of better applications.

Acknowledgments. This study was carried out as the first stage of a research entitled “A Comparative Study of Electronic Records and Metadata Schemas in Norway and Turkey”, which is supported by the YGGDRASIL scholarship, granted by the Research Council of Norway. We would like to thank the Research Council of Norway and Oslo University College for their support.

References

1. AIIM: Industry Watch (2009), <http://www.aiim.org/infonomics/still-playing-catchup-with-paper.aspx>
2. Public Record Office: E-government Policy Framework for Electronic Records Management (2001), http://www.nationalarchives.gov.uk/documents/egov_framework.pdf
3. Screene, L.: How Prepared are Public Bodies for the Implementation of the UK Freedom of Information ACT. *Records Management Journal* 15, 34–42 (2005)
4. Wamukoya, J.: Records and Archives as a Basis for Good Government: Implications and Challenges for Records Managers and Archivists in Africa. *Records Management Journal* 10, 23–33 (2000)
5. Kampffmeyer, U.: Records Management & MoReq (2009), http://www.project-consult.net/Files/20091006_InnovaDoc_Barcelona_Kff_Show.pptx
6. Waalwijk, H.: How Normal is Normalisation ? (2005), http://ec.europa.eu/transparency/archival_policy/dlm_forum/doc/14bis_waalwijk.pdf

7. MoReq: Model Requirements for the Management of Electronic Records. European Commission, Office for Official Publications of the European Communities (2001), <http://www.cornwell.co.uk/moreqdocs/moreq.pdf>
8. ISO/IEC Information Centre: Standards and Regulations (2008), http://www.standardsinfo.net/info/livelink/fetch/2000/148478/6301438/standards_regulations.html
9. Stephens, D.O.: International Standards and Best Practices in RIM. *Information Management Journal* (2000), http://findarticles.com/p/articles/mi_qa3937/is_200004/ai_n8887526/
10. Pember, M.: Sorting Out the Standards: What Every Records and Information Professional Should Know. *Records Management Journal* 16, 21–33 (2006)
11. ARMA: ARMA International (2009), <http://www.arma.org/standards/index.cfm>
12. Hofman, H.: Developments in ISO Standards for Recordkeeping. In: *DLM Forum Conference* (2005), http://ec.europa.eu/transparency/archival_policy/dlm_forum/doc/17bis_hofman.pdf
13. ISO: ISO 15489-1: Information and Documentation – Records Management. International Organization for Standardisation (2001), <http://www.iso.org>
14. Brübach, N.: International Standard for Archives and Records Management ISO 15489. In: *Proceedings of the DLM-Forum 2002, @ccess and Preservation of Electronic Information: Best Practices and Solutions*, Barcelona, May 6-8 (2002), http://ec.europa.eu/transparency/archival_policy/dlm_forum/doc/dlm-proceed2002.pdf (retrieved December 29, 2009)
15. Köhler-Krüner, H.: MoReq2 (2009), http://www.kompetencefabrikken.dk/fileadmin/ressourcer/filer/Moreq2_Denmark.pdf
16. MoReq2: Model Requirements for the Management of Electronic Records. MoReq2 Specifications. European Commission, Bruxelles-Luxembourg (2008), http://www.dlmforum.eu/index.php?option=com_jotloader&view=categories&cid=10_f56391a0c9ea9456bf24e80b514f5dda&Itemid=37&lang=en
17. MoReq2: Model Requirements for the Management of Electronic Records. MoReq2 Specifications. Appendix 9 to the Moreq2 specification: Metadata model. European Commission, Bruxelles-Luxembourg (2008b), http://www.dlmforum.eu/index.php?option=com_jotloader&view=categories&cid=10_f56391a0c9ea9456bf24e80b514f5dda&Itemid=37&lang=en
18. NOARK 5: Standard for electronic Archive (2005), http://www.arkivverket.no/noark-5/noark_5_20.pdf
19. DLM Forum: DLM Forum (2009), <http://www.dlmforum.eu/index.php>
20. Garde, J.: Essential MoReq2. DLM Forum Members Meeting. Czech Republic, Prague, April 21-22 (2009), <http://www.nacr.cz/dlm/presentations/garde.pdf>
21. Wilhelm, P.: EDRMS Standards – A Critical Evaluation of The Benefits of Superseding National Standards With European Models Focusing on TNA 2002 Replacement by MoReq2. School of Computing, Engineering and Information Sciences Northumbria University (2008), <http://www.moreq2.eu/papers-etc>
22. DLM Forum: Invitation to Tender for MoReq2 Work program 2010 (2009), http://www.dlmforum.eu/index.php?option=com_jotloader&view=categories&cid=34_2f8a6e502e2dd8ea6b7f2e7a79626230&Itemid=113&lang=en

23. Wilhelm, P.: An Evaluation of Moreq2 In The Context of National EDRMS Standard Developments in The UK and Europe. *Records Management Journal* 19, 117–133 (2009)

24. Henttonen, P.: A Comparison of MoReq and SÄHKE Metadata and Functional Requirements. *Records Management Journal* 19, 26–36 (2009)

25. TS: TS13298 Information and Documentation-Electronic Records Management. Institute of Turkish Standards (2007), <http://www.tse.org.tr>

26. Elektronik Belge Standartları Başbakanlık Genelgesi, T.C. Resmi Gazete (26938, 18 Temmuz (2008)

27. Powell, R.R.: Recent Trends in Research: A Methodological Essay. *Library and Information Science Research* 21, 91–119 (1999)

28. Cole, M., Avison, D.: The Potential of Hermeneutics in Information System Research. *European Journal of Information System* 16, 820–833 (2007)

Appendix

Table of sections and subsections analysed in MoReq2 and TS13298

MoReq2	TS13298
4. Controls and security 4.1. Access 4.2. Audit trails 4.3. Backup and recovery 4.4. Vital records	5. Access control and security 5.2. Access rights 5.3. User profiles 5.4. User roles 5.5. User groups 5.7. Creator and property rights 5.9. Audit 6. System design and administration 6.4. Storage of data 8. Document management 8.5. Storage/Backup
5. Retention and disposition 5.1. Retention and disposition schedules 5.3. Transfer, export and destruction	2. Retention 2.1. General rules 2.2. Hierarchy of classification 2.3. Criteria of retention 2.4. Retention duration 2.5. Definitions of disposition process
6. Capturing and declaring records 6.1. Capture 6.3. E-mail management 6.5. Scanning and imaging	3. Capturing of electronic records 3.1. General rules 3.4. Capturing e-mails 10. Digital scanning systems 10.3. Scanning techniques 10.4. Resolution 10.5. Color 10.9. File formats
8. Searching, retrieval and presentation 8.1. Search and retrieval 8.3. Presentation: Printing	4. ERMS usage requirements 4.2. Searching 4.4. Printing
12. Metadata scheme 12.2. General metadata requirements	11. Metadata elements 11.1. General rules

Identification of User Patterns in Social Networks by Data Mining Techniques: Facebook Case

A. Selman Bozkir¹, S. Güzin Mazman², and Ebru Akcapinar Sezer¹

¹ Hacettepe University, Department of Computer Engineering, Ankara, Turkey
selman@cs.hacettepe.edu.tr, ebru@hacettepe.edu.tr

² Hacettepe University, Department of Computer Education and Instructional Technologies,
Ankara, Turkey
s.guzin@gmail.com

Abstract. Currently, social networks such as Facebook or Twitter are getting more and more popular due to the opportunities they offer. As of November 2009, Facebook was the most popular and well known social network throughout the world with over 316 million users. Among the countries, Turkey is in third place in terms of Facebook users and half of them are younger than 25 years old (students). Turkey has 14 million Facebook members. The success of Facebook and the rich opportunities offered by social media sites lead to the creation of new web based applications for social networks and open up new frontiers. Thus, discovering the usage patterns of social media sites might be useful in taking decisions about the design and implementation of those applications as well as educational tools. Therefore, in this study, the factors affecting “Facebook usage time” and “Facebook access frequency” are revealed via various predictive data mining techniques, based on a questionnaire applied on 570 Facebook users. At the same time, the associations of the students’ opinions on the contribution of Facebook in an educational aspect are investigated by employing the association rules method.

Keywords: Social networks, decision trees, Facebook, association rules.

1 Introduction

In recent years, a rapid increase in numbers of social networks along with numbers of people using these networks has been observed. Social networks, also called social software or collaborative software, are a range of applications that augment group interactions and shared spaces for collaboration, social connections, and aggregate information exchanges in a web-based environment [1]. Similarly, [2] defined social networks as web-based services allowing individuals to 1) construct a public or semi-public profile within a bounded system, 2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system.

Millions of users have been interested in them since the introduction of social network sites (SNSs) such as MySpace, Facebook, Cyworld, Bebo, Twitter, etc. The majority of these users have integrated such sites into their daily lives. Because most

of the social network users are young individuals, many of them are university students. Therefore, these sites are considered to play an active role in the younger generation's daily life [3], [4]. On the other hand, it has been stated that social networks have a prominent educational context, and this prominence has prompted a growing number of educators to consider them to be important sites for student learning although these are not intended primarily as educational applications. Besides, it has been suggested that these social networks help users re-situate learning in an open-ended social context by providing opportunities for moving beyond the mere access to content (learning about) to the social application of knowledge in a constant process of re-orientation (learning as becoming) [5].

There have been various studies about social networks in the educational context including using social networks as a tool or utilizing them as an environment for courses [6], [7], the utility of social networks in the teaching and learning process [8], their value for communication and collaboration [9], educational usage themes of social networks (e.g. [10], [11]). However, a study in the literature about data mining analysis of social network usage has not been encountered.

As one of the most popular social networks, Facebook is considered in the present study. Facebook is defined as "a social utility that helps people share information and communicate more efficiently with their friends, family and co-workers" (facebook.com). As of November 2009, with 316 million users, Facebook is the most popular and well known social network throughout the world. Moreover, Turkey, with 14 million members, is the third country in terms of number of Facebook users and half of these members are younger than 25 years old [12].

Data mining is a process that uses a variety of data analysis tools to discover patterns and relations in data that may be used for prediction purposes. Supervised data mining techniques are used to model an output variable based on one or more input variables and these models can be used to predict or forecast future cases [13].

The purpose of the present study is to discover some usage patterns (i.e. usage time and access frequency) of Facebook users by data mining techniques. Additionally, an attempt is made to reveal the educational associations of the users. It is believed that social network based application development and educational programs can be enhanced by the findings of this study.

2 Data Mining

Data mining is the process of exploration and analysis, by automatic or semi-automatic means, of large quantities of data in order to discover useful patterns [13]. In other words, data mining is the complete process of revealing useful patterns and relationships in data by using techniques like artificial intelligence, machine learning and statistics via advanced data analysis tools. Oracle BI, SPSS Clementine, SAS Enterprise Miner and Microsoft Analysis Services are well known data mining tools in the marketplace [14].

Data mining methods are classified into two categories as predictive and descriptive. The aim of predictive methods is to make predictions on unseen cases by using

seen cases via a trained model. However, the goal of descriptive methods is discovering deep relationships, correlations and descriptive properties of data.

In this study, both of these method groups are employed by using SPSS Clementine 12. Additionally, various decision trees algorithms such as CART, CHAID and C5; artificial neural networks (ANN) and SVM (Support Vector Machine) classifiers in prediction of target variables are used. Furthermore, the variable importance feature of SPSS Clementine is used in discovering the factors affecting “Facebook usage” and “Facebook access frequency”. Likewise, the Apriori algorithm is employed in discovering frequent opinions of students on the educational benefits of Facebook usage.

2.1 Methodology

As stated previously, various data mining techniques are employed during the analyses and except one (association rules mining discovery), their prediction performances are compared. Thus, in this section, a brief information is presented about the methodologies followed.

The decision tree method is probably the most popular classification method among the data mining techniques due to the ease of use and visual interpretation capabilities. Typically, a data mining task for a decision tree is classification; for example, to identify the credit risk for each customer [15]. The main idea of a decision tree is to split the data recursively into subsets so that each subset covers more or fewer homogeneous states of the dependent variable. At each split in the tree, all independent variables are recalculated for their impact on the dependent variable. When this recursive process is stopped and the tree is in a stable state, the required decision tree is formed [15]. At this stage, new cases can be classified via the decision tree. This stage is called tree deduction. C5, Quest, CHAID [16] and CART [17] are well-known decision tree algorithms. Nevertheless, SPSS Clementine serves whole algorithms in its package. In essence, differentiations among these algorithms are mainly caused by technical capabilities and employing different splitting approaches and their functions. For instance, C5 and CHAID algorithms are designed to classify only discrete valued variables by using “gain ratio” and “gini value” splitting approaches, respectively. However, CART algorithms are designed for both classification and regression purposes.

On the other hand, in the pattern recognition literature, SVM (Support Vector Machine) is a state-of-the-art method with its powerful discriminative features in linear and non-linear classifications. Generally, SVM is designed to enlarge the boundary of any two classes in pattern space by searching for an optimal hyper plane that has maximum distance to the closest points between two classes which are termed support vectors [18]. However, SVM has support for multiclass predictions via different developed kernel functions. By the help of these kernel functions, solving the problems in upper dimensional spaces becomes possible.

ANN are systems which contain intelligence nodes arranged in layers. In essence, an ANN has an input layer, a hidden layer, and an output layer. The nodes in the hidden layer collect the inputs from the input layer into a single output value which is

passed on to the output layer. Associated with each node in the network is a weight. The weights in the network are determined in a training phase of the network using training data. The network performance is then tested on the remaining data, or hold-out sample [19].

Association rule mining is again one of the best studied descriptive mining methods since the first design and creation. Agrawal, Imelinski and Swami stated a new approach to mining association rules in 1993 and designed a new algorithm, namely Apriori, via two phases seek mechanism on itemsets and by looking their association frequencies (Romero & Ventura, 2007). In the second stage of this study, the analyses are performed by using the algorithm Apriori. In association rules, mining analyzing, support, rule support, confidence and lift values are the important parameters in the usefulness evaluation of rules. In this study, lift and support values are considered.

Table 1. Variable names and available answers in the first part of the poll

Variable name	Type	Available answers and related distributions
Sex	Discrete	Male (50%) / Female (50%)
Age	Discrete	18-25 (74.1%) / 26-35 (20.53%) / 36-40 (3.86%) / 41 and above (1.4%)
Frequency of access to Facebook	Discrete	Once a year (0.18%) / Once a month (2.98%) / Several times a week (25.26%) / Once a day (22.81%) / Several times a day (48.77%)
Facebook usage time	Discrete	Less than 15 mins. (32.28%) / Half an hour (39.82%) / 1 hour (14.39%) / 1-3 hours (8.6%) / More than 3 hours (4.74%)
Education level	Discrete	High School (5.96%) / Bachelor (70.35%) / Master (23.16%)
Membership in any group	Discrete	Yes (99.82%) / No (0.18%)
Membership in student groups	Discrete	Yes (86.49%) / No (13.51%)
Membership in common interest groups	Discrete	Yes (77.54.5) / No (22.46%)
Membership in internet & tech groups	Discrete	Yes (27.02%) / No (72.98%)
Membership in organizations	Discrete	Yes (61.93%) / No (38.07%)

3 Data

Data was collected from 570 active Turkish Facebook users (students) with an online poll. This online poll consisted of two sections. In the first section, demographic characteristics of Facebook users and their frequency of Facebook usage, length of time spent on Facebook, and memberships in Facebook groups were collected. In the second section, a 10-point Likert scale with 11 opinions were asked, the answers ranging from 1 (strongly disagree) to 10 (strongly agree), like “Facebook contributes to communication between classmates”, “It’s useful for assigning tasks in classes and

homework assignments". Thus members' views of Facebook in relation to its educational usage were sought.

The variable names of the first part and available answers are given in Table 1. Although the initial dataset size was larger than 570 people, during the data cleaning and transforming steps, 13 people were removed due to the absence of sufficient information. Therefore, the final dataset comprised 570 people. In the dataset, male and female participants are almost equal and more than 400 applicants are in the 18-25 age range. Furthermore, almost all students are at either undergraduate or graduate level.

4 Application of Data Mining

To discover important factors that affect Facebook usage time and access frequency to Facebook, CART, CHAID, C5, artificial neural network and SVM algorithms, which are built in to SPSS Clementine 12, were employed on the dataset at hand (see Fig. 1). The overall data is partitioned as 80% training and 20% testing, respectively. Training and test datasets are selected randomly. As the dataset consists of discrete valued variables, the true and false prediction rates are listed.

According to the results (see Table 2), SVM achieves the most accurate predictions for two target variables. Therefore, it is considered that the variable importance results of SVM are the most accurate predictions. As can be seen in Fig. 2, sex, education level, membership in a group and membership in any common interest groups are the most important factors affecting *Facebook usage time*. Sex plays a crucial role in Facebook usage time with 68%. Again, it can be clearly seen that age, membership in student groups and usage time variables are the most important factors affecting *access frequency to Facebook*. The effect of age is more than 80% in *access frequency*.

Table 2. Applied algorithms and prediction results

Target variable - Applied algorithm	True classification	False classification
Facebook usage – SVM	62.63 %	37.37 %
Facebook usage – ANN	47.72 %	52.28 %
Facebook usage – C5	47.54 %	52.46 %
Facebook usage – CART	43.68 %	56.32 %
Facebook usage – CHAID	41.40 %	58.60 %
Access frequency to Facebook – SVM	69.65 %	30.35 %
Access frequency to Facebook – C5	55.79 %	44.21 %
Access frequency to Facebook – CART	52.81 %	47.19 %
Access frequency to Facebook – CHAID	50.35 %	49.65 %
Access frequency to Facebook – ANN	48.77 %	51.23 %

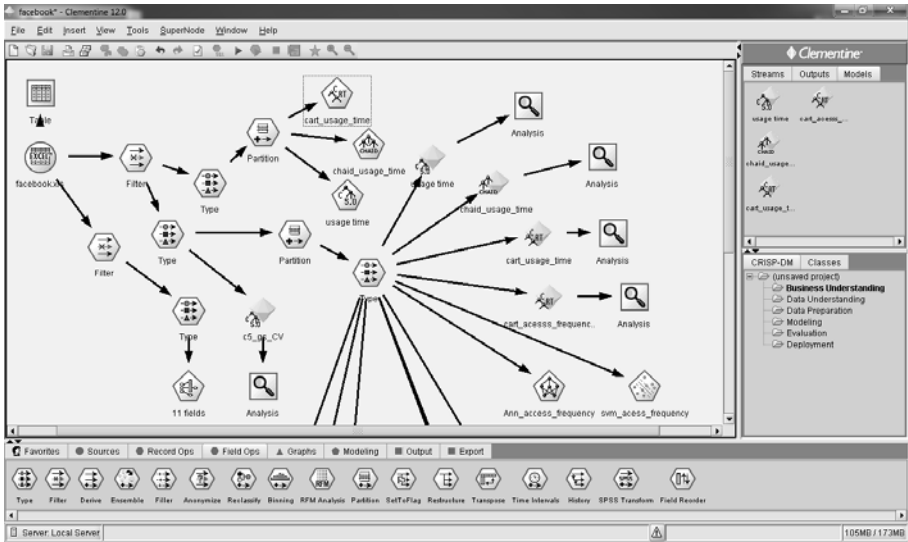


Fig. 1. Applying data mining methods in Clementine 12

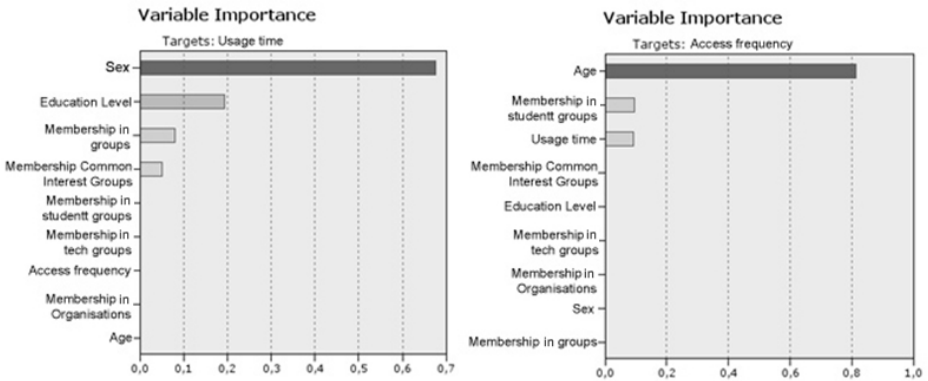


Fig. 2. Variable importance values of two target variables

On the other hand, in the association rules mining study, the association of the student ideas on Facebook and its educational benefits has been investigated. To achieve this, the well-known Apriori algorithm was run with 5% support and 15% confidence parameters.

As can be seen in Table 3, some interesting rules are listed sorted by lift values. Lift value shows the usefulness and attractiveness of a rule. Therefore, the rules which have lift values higher than 1 should be considered carefully for educational purposes.

Table 3. A sample subset of discovered association rules

Antecedent	Consequent	Confidence	Support	Lift
“It contributes to the communication between teacher & student” = 7	“It’s useful at accessing the rich learning resources” = 9	11.32%	1.03%	3.0
“Facebook contributes to communication between classmates” = 4	“It’s useful for executing the group tasks” = 2	12.83%	0.87%	2.9
“Facebook contributes to communication between classmates” = 8 and “It contributes to communication between teacher & student” = 8	“It contributes to transferring course materials and resources” = 6	20.69%	1.05%	2.7
“Facebook contributes to communication between classmates.” = 6	“It’s useful at providing rich multimedia contents in teaching” = 3	20.51%	1.40%	2.65
“It contributes to the communication between teacher & student” = 3	“Facebook contributes to communication between classmates” = 6	17.02%	1.40%	2.48
“Facebook contributes to communication between classmates” = 7 and “It contributes to communication between teacher & student” = 5	“It contributes to dissemination of announcements of lectures & classes” = 2	12.5%	0.52%	2.03

5 Discussion and Conclusion

This study tried to discover the factors affecting access frequency and usage time of Facebook by various decision tree algorithms, ANN and state-of-the-art algorithm SVM. According to the results, SVM exhibits the most accurate results due to the nature of the dataset at hand. It is believed that the prediction capabilities can be enhanced by using more training data. On the other hand, the associations of the student ideas were explored by employing the Apriori algorithm and, as can be seen from the results obtained, the contribution of Facebook to communication between classmates is more than to communication between students and teachers. Moreover, the students who hold these views believe that Facebook is a good medium for accessing rich resources. More of these types of rules can be revealed by using the Apriori algorithm and the use of social network sites for educational ends can be reformed in the light of these rules.

If the increasing trend in social network sites usage is considered, the importance of applications and approaches related to social networks can be easily understood. Targeting specific ages or sex may strategically affect the success of developed applications. As a consequence, data mining methods can be successfully employed on social network usage data.

References

1. Bartlett-Bragg, A.: Reflections on Pedagogy: Reframing Practice to Foster Informal Learning with Social Software (2006), <http://www.dream.sdu.dk/uploads/files/Anne%20Bartlett-Bragg.pdf>
2. boyd, D.M., Ellison, N.B.: Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication* 13, 210–230 (2007)
3. Lenhart, M.: Adults and Social Network Websites. Pew Internet & American Life Project Report (2009), http://www.pewinternet.org/pdfs/PIP_Adult_social_networking_data_memo_FINAL.pdf
4. Bumgarner, B.A.: You Have Been Poked: Exploring the Uses and Gratifications of Facebook Among Emerging Adults. *First Monday*, 22 (2007), <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/2026/1897>
5. Mejias, U.: Nomad's Guide to Learning and Social Software (2005), http://knowledgetree.flexiblelearning.net.au/edition07/download/la_mejias.pdf
6. English, R., Duncan-Howell, J.: Facebook Goes to College: Using Social Networking Tools to Support Students Undertaking Teaching Practicum. *Journal of Online Learning and Teaching* 4, 596–601 (2008)
7. Lockyer, L., Patterson, J.: Integrating Social Networking Technologies in Education: A Case Study of a Formal Learning Environment. In: *Proceedings of 8th IEEE International Conference on Advanced Learning Technologies, Spain*, pp. 529–533 (2008)
8. Ajjan, H., Hartshorne, R.: Investigating Faculty Decisions to Adopt Web 2.0 Technologies: Theory and Empirical Tests. *The Internet and Higher Education* 11, 71–80 (2008)
9. Saunders, S.: The Role of Social Networking Sites in Teacher Education Programs: A Qualitative Exploration. In: McFerrin, K., et al. (eds.) *Proceedings of Society for Information Technology and Teacher Education International Conference*, pp. 2223–2228. AACE, Chesapeake (2008)
10. Mazman, S.G., Usluel, Y.K.: Adoption Process of Social Network and Their Usage in Educational Context. Master Thesis. The Institute for Graduate Studies in Science and Engineering, Hacettepe University, Ankara (2009)
11. Selwyn, N.: Web 2.0 Applications as Alternative Environments for Informal Learning - A Critical Review. *Alternative Learning Environments in Practice: Using ICT to Change Impact and Outcomes*, OECD-KERIS Expert Meeting (2007)
12. Check Facebook (2009), <http://www.checkfacebook.com>
13. Berry, M., Linoff, G.: *Mastering Data Mining: The Art and Science of Customer Relationship Management*. John Wiley & Sons, Chichester (2000)
14. Bozkir, A.S., Gök, B., Sezer, E.: İnternetin Eğitimsel Amaçlar için Kullanımını Etkileyen Faktörlerin Veri Madenciliği Yöntemleriyle Tespiti. In: *Bilimde Modern Yöntemler Sempozyumu*, pp. 833–842. Eskişehir (2008)
15. Tang, Z., MacLennan, J.: *Data Mining with SQL Server 2005*. John Wiley & Sons, Indiana (2005)
16. Kass, G.V.: An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Applied Statistics* 29, 119–127 (1980)
17. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, P.J.: *Classification and Regression Trees*. Wadsworth International Group, Belmont (1984)

18. Liu, J., Wang, Z., Xiao, X.: A Hybrid SVM/DDBHMM Decision Fusion Modeling for Robust Continuous Digital Speech Recognition. *Pattern Recognition Letters* 28, 912–920 (2007)
19. Fuller, C.M., Piros, D.P., Wilson, R.L.: Decision Support for Determining Veracity via Linguistic-Based Cues. *Decision Support Systems* 46, 697–703 (2009)
20. Romero, C., Ventura, S.: Educational Data Mining: A survey from 1995 to 2005. *Expert Systems with Applications* 33, 135–146 (2007)

Digital Rights Challenges in a World of Technological Convergence

John N. Gathegi

University of South Florida, School of Library and Information Science, 4202 E. Fowler Ave., CIS1040 Tampa, FL. 33620, USA
jgathegi@cas.usf.edu

Abstract. The convergence of communication, caused in part by the convergence of media and digital content, is a continuing phenomenon. Many intellectual property challenges loom in this environment. This paper focuses on the situation in the United States. The peculiar features inherent in digital content that exacerbate the intellectual property problem, such as non-permanent, multiple, and heterogeneous media are discussed. A few US cases that illustrate some of the problems in this area are also examined. The paper concludes by looking at the multiple goals of digital content collections and the problem of intellectual property.

Keywords: Digital rights, intellectual property, digital content.

1 Introduction

This paper examines the continuing phenomenon of convergence of communication, caused in part by the convergence of media and digital content. We review some of the intellectual property challenges that loom in this environment, with an emphasis on the situation in the United States. We also discuss some of the peculiar features inherent in digital content that exacerbate the intellectual property problem, such as non-permanent, multiple, and heterogeneous media. Also, we examine a few US cases that illustrate some of the problems in this area. We conclude by looking at the multiple goals of digital content collections and the problem of intellectual property.

2 Dynamism and the Problem of Non-permanence

Unlike its print counterpart, digital content has some unique features that pose legal challenges in both management and development. In this paper, we shall examine three such features: digital content may have more than one media format, may suffer issues of non-permanence, and is dynamic.

Dynamism is a key characteristic of digital content. Corrections and modifications are constantly being made to specific files and databases, and items are constantly being added and as the need arises. Thus, a file today may not be exactly the file one

looked at yesterday. The preservation of digital content and vouching for its integrity then become pressing questions.

There are many problems that come with preservation efforts. Chief among these is how to ensure the efforts do not result in the infringement, not just by avoiding unauthorized exercise of the authors' exclusive rights, but also by determining the scope of copyright protection, especially where availability to content and agreement with copyright holders is desired. Whether the digital content manager still has the necessary rights to the e-content is a crucial question. The dynamism and non-permanence of digital content may raise issues of ethics, privacy, and confidentiality, especially where health and personal data is involved [1].

Lavoie and Dempsey note that despite perceptions to the contrary, "digital information is in fact fragile and at risk" [2]. Some digital files can be rendered corrupt and unreadable, due to changes in technology. Uncertainty with information preservation increases with the length of the time frame required for future access. Changes in metadata content, data definitions and format are some of the challenges confronting the digital content manager [1]. This is especially a problem because many formats have protection and are continually evolving, more complex versions with newer features and functions are developed. This sometimes results in earlier versions being 'orphaned' [3]. Contractual limits to access by a proprietary owner or a proprietary owner who goes out of business can also present legal access problems [4].

Migration of data, both in terms of software and hardware, is one way of handling format changes in digital preservation. Sometimes this will involve re-arranging the sequence of structural and data elements [3]. Migrating data usually involves copying information, which infringes the exclusive right of the author to reproduce. Re-arranging the structural and data elements may also trample on the exclusive right of the author to make derivative copies. Thus, it is conceivable that one might need to seek permission from the copyright holder to migrate the information. Whether a file conversion would violate the Digital Millennium Copyright Act is an issue in the United States, as is the question of whether a migrated file is the same as the original file for evidentiary purposes.

3 The Complexity of Multi-media Content

Digital content may also contain a mixture of different media formats, including text, sound, graphics, video, and a variety of other file formats in addition to being dynamic and raising the problem of non-permanency.

Electronic books, or e-books, are good examples of multimedia digital content. For instance, one might find an article about a country, a video about parts of the country, and a sound file of examples of music from the country in an e-book.

E-books can be accessed through a central server, but are often proprietary devices. Just like with any other digital content, e-book collections are equally susceptible to easy copying. In order to protect the exclusive rights of copyright owners, e-books use digital rights management technology to control access to copyright protected content. Not all e-books are so protected, however. A digital manager can link to much e-book

content that is available free of copyright protection [5]. By negotiating intellectual property rights with publishers, e-book aggregators are able to provide 24/7 access to content hosted on their servers. Legal issues are less of a problem here, because aggregators usually implement their own digital rights management technology [6].

It is important that the digital collection manager understand limitations on access that come with digital rights management systems, as well as the variations in pricing models, which vary from outright purchase to limited term leases. Leases can include terms that restrict certain uses such as printing and downloads, as well as what content can be accessed and the number of person that can have access at the same time.

The fact that some media formats are covered by rules specific to the media, sound files, for example, complicates legal issues in this area. As discussed above, copyright infringement may be triggered by conversion from one format to another. In the United States, the date on which the sound recording was first fixed may determine what legal protection exists. Prior to February 15, 1972, for example, federal law did not provide copyright protection to sound recordings. This situation was corrected by the Sound Recording Amendment Act of 1971, which for the first time provided copyright protection to such works [7]. Until February 15, 2067, however, pre-1972 works may be protected against unfair competition or misappropriation by state criminal law statutes or common law [8].

3.1 Document Icons and Page Thumbnails

Also presenting new legal issues are other newer versions of familiar formats, such as document icons and page thumbnails. Document icons can be described as visual representations of documents in a reduced format [9]. They usually contain format or genre information about a document, whether, for example, the document is a pdf, a web page, or simply a file folder. Similarly, page thumbnails are represented as small images of a page that can be enlarged for viewing by the user, and generally have a lower resolution.

Thumbnails have an impact on two rights exclusive to the copyright holder. While crawling, search engines make copies of the images they encounter. Thus, they may be in violation of the author's exclusive right to make reproductions of a work [10]. When they show the thumbnails to the user, search engines may in some cases also be in violation of the right to public display, which is an author's exclusive right [10]. But such a claim can sometimes be refuted by relying on a limitation of the author's exclusive rights provided by the Fair Use doctrine, as the *Kelly v. Arriba* [11] and *Perfect 10 v. Google* [12] cases below do.

In the *Kelly* case, a visual search engine operator had built up a database by copying images from web sites. It would then reduce these images into thumbnails that a user could enlarge by simply clicking on them. The operator then displayed these thumbnails on its website, and licensed others to do the same on their websites. A photographer sued the operator for displaying the photographer's copyrighted images. Finding that the character and purpose of the operators use of the images was "significantly transformative and the use did not harm the market for or value of [the photographer's] works," the lower court ruled that the use of the thumbnails was fair

use [11]. That ruling was affirmed on appeal by the 9th Circuit Court of the US Court of Appeals.

A website operator/publisher got a chance to sue Internet search engines in *Perfect 10 v. Google* [12]. Perfect 10 had invested heavily in developing a brand name for a web site and a magazine in which it published adult photographs. The search engines operators Google and Amazon, whose search engines following a search string query would retrieve thumbnail images, included Perfect 10's images retrieved from Perfect 10's web site. Perfect 10 sued both search engine operators. In response to Perfect 10's motion for an injunction, trial judge explained:

The principal two-part issue in this case arises out of the increasingly recurring conflict between intellectual property rights on the one hand and the dazzling capacity of internet technology to assemble, organize, store, access, and display intellectual property "content" on the other hand. That issue, in a nutshell, is: does a search engine infringe copyrighted images when it displays them on an "image search" function in the form of "thumbnails...?" [12]

The trial judge issued the preliminary injunction against Google's creating and displaying Perfect 10's images, because he believed there was a likelihood of success in Perfect 10's claim of Google's direct infringement of its copyrighted images. According to the trial judge, this case was different from *Kelly v. Arriba* because Perfect 10 conceivably had a market for downloading adult images thumbnails into mobile phones, and that Google's action would harm that market. On appeal, the 9th Circuit threw out that argument, ruling, as it had done in *Kelly*, that because they did not detract from the economic value of the images, Google's use of the thumbnails were fair use, opening the door for Google's continued use of Perfect 10's thumbnails [13].

4 The Legal Complexities of Multiple, Heterogeneous Content

Digital collections have multiple, heterogeneous content. There are two ways of looking at this characteristic. One is by looking at the different types of digital collections, and the other is by looking at the different goals of digital collections, while acknowledging that digital collections often has multi-type content designed for multi-type goals.

One ubiquitous example is the Internet itself. From one vantage point, this can be considered as one giant digital collection; something we might call a "meta-collection." Libraries and individuals, usually through linking and bookmarking, take subsets of this "meta-collection" to organize their own special collections. In Europe, unlike in the US, linking has generated some legal challenge; Bookmarking on the other hand has seen no such challenge.

Digital collections may include commercial databases, which bring with them their own issues regarding copyright protection and licensing. As far as data sets (which may be a type of commercial database) go, there is more emphasis on licensing issues as opposed to copyright.

A special focus on collective works and compilations is appropriate at this point. In the US, a compilation is defined as “a work formed by the collection and assembling of preexisting materials or of data that are selected, coordinated, or arranged in such a way that the resulting work as a whole constitutes an original work of authorship. The term ‘compilation’ includes collective works.” [14].

The US copyright law specifically mentions that compilations are included as a subject matter of copyright, but with a caveat: only what the author has contributed is protected, the underlying or pre-existing material is not, and neither unlawfully-used pre-existing material [15].

Individual components of a collective work can generally have their own independent copyright protection, but material that is not necessarily copyrightable may be present in compilations [16]. Copyright in the collective work as a whole is distinct from copyright protection of the separate contributions. Copyright in a collective work entitles the owner only to “the privilege of reproducing and distributing the contribution as part of that particular collective work, any revision of that collective work, and any later collective work in the same series [17].” This doctrine was reaffirmed in *New York Times Co., Inc. v. Tasini* [18], a case that looked at whether a collective work copyright owner could republish all or a part of a database in spite of an assertion of copyright infringement from a contributor to the work. In this case, a newspaper publisher had made articles authored by free-lance authors available in electronic databases. The authors sued the newspaper. Section 201(c) of Title 17 offers a privilege, which the newspaper asserted:

Copyright in each separate contribution to a collective work is distinct from copyright in the collective work as a whole, and vests initially in the author of the contribution. In the absence of an express transfer of the copyright or of any rights under it, the owner of copyright in the collective work is presumed to have acquired only the privilege of reproducing and distributing the contribution as part of that particular collective work, any revision of that collective work, and any later collective work in the same series. [17]

Rejecting the newspaper’s reliance on the privilege offered by section 201(c), The Supreme Court looked at how the user would perceive the articles as presented in the database. Under the right circumstances, however, the privilege continues to be available. For example, the National Geographic publisher compiled past issues of the magazine and digitized them making transforming them into a database dubbed the Complete National Geographic, that electronically searchable. When freelance photographers and authors sued objecting to their work being used in this new media, a second circuit court affirmed the granting of summary judgment to the publisher [19]. The Court concluded that the digital Complete National Geographic was a new version of the print National Geographic Magazine, and that the original context of the magazine was present in the new version. Users of the database in *Tasini* were, however, unable to view the underlying works in their original context, unlike in the National Geographic Case.

4.1 Different Goals

Digital collections are heterogeneous, are in multiple formats, and have different goals. Preservation is one of the goals of digitization. US copyright law grants some institutions a legal privilege to exercise the author's exclusive right to reproduce, for the purposes of preservation. Preservation and conservation needs in the US are taken care of by Section 108 of the U.S. copyright code [20]. Along with copyright, evidence is among other legal issues that are likely to emerge in this area.

When considering preservation for evidentiary purposes, the issue of non-permanence becomes critical. It is telling that Electronically Stored Information (ESI) was added in December 2006 as a new category of evidence in the U.S. Federal Rules of Civil Procedure to work within the existing rules for production of "documents" during discovery. Parties in litigation have to provide each other with: "a copy--or a description by category and location--of all documents, electronically stored information, and tangible things that the disclosing party has in its possession, custody, or control and may use to support its claims or defenses..." [21]. The version of the electronically stored information to be produced is not specified, but Rule 26(f) obliges the parties to meet and "...discuss any issues about preserving discoverable information" and "any issues about disclosure or discovery of electronically stored information, including the form or forms in which it should be produced" [21].

Authenticity in evidentiary terms is an issue that is closely related to non-permanence. Tampering with or corrupting digital information is not an uncommon occurrence. To preserve authenticity, it may be necessary to employ authenticating tools like encryption, digital signatures, and version control, among others [22].

In conclusion, we have to emphasize that access and preservation are much intertwined. Access to information can be for many different reasons, such as research, safeguarding culture, or even entertainment. The issue of copyright protection is always present, and is complicated by the fact that there are different duration terms for different works, depending on the creation or publishing date. In preservation, licensing for access is likely to be the most frequently occurring issue. However, the scope of this paper unfortunately precludes a discussion of access as defined by use.

References

1. Anderson, W.L.: Some Challenges and Issues in Managing, and Preserving Access to, Long-Lived Collections of Digital Scientific and Technical Data. *Data Science Journal* 3, 191-201 (2004)
2. Lavoie, B., Dempsey, L.: Thirteen Ways of Looking at...Preservation. *D-Lib Magazine* 10 (2004), <http://dlib.org/dlib/july04/lavoie/07lavoie.html>
3. Arms, C.: Keeping Memory Alive: Practices for Preserving Digital Content at the National Digital Library Program of the Library of Congress. *RLG DigiNews* 4 (2000), <http://webdoc.gwdg.de/edoc/aw/rlgdn/preserv/diginews/diginews4-3.html#feature1>
4. Barnes, I.: The Preservation of Word Processing Documents. Canberra: Australian Partnership for Sustainable Repositories (2006), http://www.apsr.edu.au/publications/word_processing_preservation.pdf

5. Esposito, J.: The Processed Book. *First Monday* 8 (2005),(Updated from the 2003 article), <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1038/959>
6. Garrod, P.: Ebooks in UK Libraries: Where are we Now? *Ariadne* 37 (2003), <http://www.ariadne.ac.uk/issue37/garrod>
7. Besek, J.M.: Copyright Issues Relevant to Digital Preservation and Dissemination of Pre-1972 Commercial Sound Recordings. In: Council on Library and Information Resources and Library of Congress, Washington, DC (2005)
8. 17 USC 301(c) provides: "With respect to sound recordings fixed before February 15, any rights or remedies under the common law or statutes of any State shall not be annulled or limited by [Title 17] until February 15 (2067)
9. Janssen, W.C.: Document Icons and Page Thumbnails: Issues in Construction of Document Thumbnails for Page-Image Digital Libraries. Palo Alto Research Center, Palo Alto (2004)
10. 17 USC 107 (c)
11. *Kelly v. Arriba*, 280 F.3d 934 (9th Cir. 2002)
12. *Perfect 10 v. Google*, 416 F. Supp. 2d 828 (C.D. Cal. 2006)
13. *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146 (9th Cir. 2007)
14. 17 USC 101 (Title 17 of the United States Code, Section 101)
15. 17 USC 102, 103
16. *Roy Export Co. Establishment of Vaduz, Liechtenstein, Black Inc., A. G. v. Columbia Broadcasting System, Inc.*, 503 F. Supp. 1137 (S.D. N.Y. 1980), judgment aff'd, 672 F.2d 1095 (2d Cir. 1982)
17. USC 201(c)
18. *New York Times Co., Inc. v. Tasini*, 533 U.S. 483 (2001)
19. *Faulkner v. National Geographic Enterprises Inc.*, 409 F.3d 26 (2d Cir. 2005)
20. 17 USC 108
21. Federal Rules of Civil Procedure 26(a)(1)(A)(ii)
22. Germain, C.M.: Legal Information Management in a Global and Digital Age: Revolution and Tradition. *International Journal of Legal Information* 35, 134–163 (2007)

Copyright Issues Related to the Digitization of Cultural Heritage in Croatia

Aleksandra Horvat and Daniela Zivkovic

University of Zagreb, Faculty of Humanities and Social Sciences, Ivana Lucica 3, Zagreb,
Croatia
{ahorvat,dzivkovi}@ffzg.hr

Abstract. A proliferation of EU programmes and action plans on digitization proves that there is a political will to digitize and make available to the public the rich cultural heritage of Europe. The article tracks the development of these ideas and actions and focuses on one of the obstacles to be overcome—the lack of consistency in approaches to intellectual property rights. The aim of the investigation described in the paper has been to find out how copyright issues related to the digitization performed by public institutions such as libraries and archives have been dealt with in Croatia. Information collected from the national electronic portal Croatian Cultural Heritage was further enriched by interviews held with the persons responsible for digitization projects in four major public institutions in the country. A lack of suitable registries or databases with data on national authors has been noted. Institutions show different approaches to digital copies they produce; some consider themselves to be the publishers and rights holders of new digital editions of works, while the others take digitization primarily as a means of protection of originals. As a rule institutions have a policy of copyright management, even if a rudimentary one. All institutions provide copyright information, sometimes with a copyright disclaimer. They use watermarking as a technical protection measure. Digitized materials are provided at no cost on the Internet for private use and research. Use of materials for commercial purposes has to be paid for and the institutions look upon it as a source of revenue to fund further digitization activities. The institutions' policy toward derivative works, i.e. offering the opportunity to users to create their own content, is yet to be determined.

Keywords: Digitization, copyright, orphan works, user-created content, cultural heritage, Croatia, EU digitization policy.

1 Introduction

The fact that mass digitization is today not only technically feasible but has become a political and social issue of great importance has certainly contributed to the resurgence of discussions about copyright. In the last decade the European Union has adopted a number of official documents in the form of reports, recommendations and resolutions, etc. on digitization. The European Commission first called for measures to stimulate the development and use of digital content in the eEurope 2002 Action

Plan [1]. The eEurope Action Plan was endorsed by EU Member States in June 2000, encouraging the Commission and the Member States to “create a coordination mechanism for digitization programs across member States”.

As a first step towards EU-wide coordination of digitization programs and policies, the European Commission organized an expert meeting with representatives from all member states in Lund in (Sweden) April 2001. The conclusions and recommendations derived from this meeting are known as the Lund principles [2]. Among various obstacles they recognize intellectual property rights and state that various stakeholders in the digitized content have different legitimate interests which should be balanced. Solutions for handling and managing rights need to be understood and applied by the cultural sector if the economic value of the content is to be realized in a sustainable way. They “established priorities to add value to digitization activities in ways that would be sustainable over time”. The Lund Action Plan [3] recommended actions for the period until 2005 and was implemented by a group of national representatives. The Group, appointed by the national Authorities for Culture in the EU, met in Parma in 2003 and agreed on a set of principles enumerated in the Charter of Parma [4]. Article 4 of the Charter recognizes not only intellectual property rights, but also privacy of individuals, as obstacles and encourages the adoption of all available technical and legal instruments to improve accessibility and overcome legislative and normative barriers. It also encourages a dialogue between the cultural and scientific sectors, intellectual property experts, companies implementing Digital Rights Management solutions and the content industries.

In 2005 the successor of the Lund Action Plan was presented under the title Dynamic Action Plan for the EU co-ordination of digitization of cultural and scientific content [5]. It builds upon the previous plan. Let us mention just two of the six objectives: strengthening co-ordination and forging stronger links between Member States digitization initiatives, EU networks and projects, and improving online access to European cultural heritage.

The proliferation of official documents concerning digitization in the EU shows clearly that digitization of various kinds of materials including books, newspapers, films, photographs, and maps held in libraries, archives and other public institutions throughout Europe has been strongly encouraged and supported not only as a means of preservation of the vast European cultural heritage and the instrument by which the heritage can be made available to a wider audience than ever before, but also as an incentive to the further development of the creative industry and the content market [6].

Analyzing the results of those efforts at the European level the following obstacles to mass digitization have been observed: cost, use of inappropriate technologies and inadequate standards, lack of synergy between cultural and technological programmes and *the lack of consistency in approaches to intellectual property rights* [7].

2 Copyright as an Obstacle

The basic principle of the EU Directive on the harmonisation of certain aspects of copyright and related rights in the Information society from 2001 was to provide

rights holders with a high level of protection by adapting the exclusive rights to the online environment. It also introduced an exhaustive list of exceptions to copyright protection. This list does not allow EU states to maintain or introduce exceptions which are not listed. The three-step test from the Berne Convention for the Protection of Literary and Artistic Works [8] has become a benchmark for all copyright limitations. In 2008 the EC Commission issued a Green Paper, Copyright in the Knowledge Economy [9], with the aim of improving the free movement of information and initiating a debate on the dissemination of creative works in the digital environment, not being limited to science and education. Its aiming at all materials which have value in enhancing knowledge is of great importance to digitization projects of cultural heritage. The issues connected to copyright, such as present exceptions and limitations to copyright protection of use to libraries and archives, exceptions for handicapped persons and use of works in education were raised. The possible directions of further development with regard to exceptions were outlined: contractual arrangements between rights holders and users for the implementation of existing copyright exceptions and contractual arrangements between rights holders and users on other aspects not covered by copyright exceptions. One of the solutions would be that certain categories of exceptions might be made mandatory to ensure more legal certainty and better protection of beneficiaries of exceptions. The Green Paper focuses on the exceptions to copyright which are most relevant for the dissemination of knowledge. In the first place, exceptions for libraries and archives are mentioned. Under the current legal framework, libraries and archives do not enjoy a blanket exception from the right of reproduction. Reproductions are only allowed in specific cases such as preservation of works contained in the library holdings [10] and that covers the preservation function of digitization. But libraries are interested in making these digital copies accessible online. Unfortunately, publicly accessible libraries, educational establishments, museums and archives do not benefit from the exception to the communication to the public right and the making available right for the purpose of research or private study by means of dedicated terminals located on the premises of such establishments [10] when making digitized works available online. Other issues connected to copyright, such as present exceptions and limitations to copyright protection of use to libraries and archives, exceptions for handicapped persons and use of works in education were raised and the problems related to orphan works were enumerated.

International organizations, such as IFLA and eIFL which support libraries, have quickly realized the possible obstacles to libraries wishing to digitize their material. As early as in 2002 IFLA had published its Guidelines for digitization projects for collections and holdings in the public domain, particularly those held by libraries and archives. IFLA suggests criteria for selection, technical requirements and implementation, legal aspects, budgeting, development and maintenance of web interfaces, preservation of digital content and project management [11]. The Guidelines identify and discuss the key issues involved in conceptualization, planning and implementation with recommendations for best practices to be followed. They are aimed at decision makers, library and archive managers particularly in developing countries and are a part of the UNESCO strategy of knowledge for all.

One chapter of the Guidelines addresses legal aspects of digitization. Institutions planning a digitization project must investigate the copyright situation for each item,

take steps to prevent unauthorized changes in the digital files created during the project, ensure that the project complies with appropriate local legal deposit legislation and put in place a clear policy on access to and use of images within the completed digital project, including, if necessary, provision of copyright disclaimer forms. A recommendation is made to the institutions that the first issue to address in a digitization project or programme is the legal conditions for making digital copies.

Association of European Research Libraries (LIBER) has recently formulated its view of copyright in the digital environment stating strongly that access to and sharing of information and not copyright is the basis for scholarly activity supporting the idea that equitable access to information is vital for social, educational, cultural, democratic and economic well-being of people in Europe. The specific situation with copyright in Europe is that legislation exists at both European and national levels. LIBER advocates a stronger and unified European copyright regime, which balances the rights of rights holders and those of users. In order to ensure access to a broad range of information resources, publicly-financed digital content should remain in the public domain. Though public-private partnerships have an important role in helping achieve the European Commission's strategy for digitization, public-private partnerships contracted in order to digitize content should be licensed for a maximum of 10 years. Yet, charges for the delivery of high quality images, sound and video recordings and print-on-demand materials are allowed. The future European copyright rules should stimulate the use and re-use of content as well as development of new models of fair remuneration for the rights holders. Contractual overriding of copyright law and its exceptions should be prevented.

In the beginning of this decade pressure seems to have been put to WIPO (World Intellectual Property Organization), as the UN organization in charge of copyright issues at the international level, to change the present copyright policy at least when exceptions and limitations to copyright are concerned [12]. Experts gathered at recent sessions of the WIPO Standing Committee on Copyright and Related Rights (SCCR), together with representatives from library and information associations, such as IFLA/CLM, eIFL, etc. have tried to analyze and explore the existing limitations and exceptions including those which are considered to be of use to libraries and archives [13]. The overall opinion seems to prevail that present copyright laws in the European countries have not been amended to enable libraries and archives to operate efficiently and adequately in the global digital environment. It seems also that present copyright exceptions and limitations are increasingly undermined by licensing practices which sometimes overrule the national copyright laws.

3 Copyright Clearance

Digitization is an act of both reproduction and communication to the public and therefore requires permission of the author/rights owner or her/his representative, if the work is copyrighted. A public institution engaged in digitization first has to identify the author and determine if the work is still protected (in the EU the work is protected during an author's lifetime and the 70 years following his/her death). Also, the copyright for a protected work could have been transferred to another physical or legal

person. Tracing the rights holders can be a lengthy and burdensome process, even more so if a library wishes to digitize various kinds of materials, such as phonograms, artefacts, letters, etc. Libraries are also interested in the material created by local people or amateurs. In a paper already more than a decade old on copyright clearance and digitization in higher education, Bide, Oppenheim and Ramsden [14] discuss the need for a "one-stop shop" for the clearance of digitization rights. The authors believe that a central agency for rights clearance is highly desirable, although they find a possible disadvantage of a "one-stop shop" method of paying for the rights in the impossibility of negotiability of rates. A contract or a license should be as brief and straightforward as possible and unambiguous in the wording. Unfortunately thirteen years later the one-stop shop for the clearance of rights is still to be wished for.

4 Orphan Works

The main problems with the present copyright regime seem to originate in the inability of institutions such as libraries and archives to trace and locate rights holders in many cases. The continental law system, as distinct from the Anglo-American, has never required the registration of copyright, and now when libraries wish to find authors and other rights holders, they are quite often at a loss. There is no agency which could provide data on authors and other rights holders not only at global level, but often within a country. It is a true paradox that libraries which have acquired such rich expertise in organizing catalogues and bibliographies and determining various kinds of authorship have now been confronted with a lack of adequate data on authors.

In Europe, but also in the US and Canada, a lot of work has been put into trying to solve the so-called orphan works problems, but adequate solutions are still sought. Libraries and archives hesitate to select orphan works for digitization since their copyright status is difficult to determine. Ephemeral materials such as photographs, postcards, posters, programmes of local events, etc. belong to that category of works, and that kind of content seems to be particularly interesting for the institutions that keep them because of its historical or cultural value for the community in which the material had been created. They witness the historical development of a town, village, region and are interesting for researchers of local history, historians, teachers and the public at large. Unfortunately libraries and archives keep little or no information on the creators of such material. According to some estimates several millions of items belong to the category of orphan works [15]. According to the British Library around 40% of all creative works are orphan works [15]. Research done in the UK has shown that 5-10% of each public institution's collections would probably never be digitized because the rights cannot be cleared. Naturally, it is the 20th century works that are at risk of not being digitized and available to the public. In the i2010 Digital Libraries the experts state that clarification and transparency in the copyright status of a work is an essential element in a number of areas, including the European Digital Initiative

(EDI). The EDI was launched by the European Commission in 2005 to provide a common multilingual access point to Europe's digital cultural heritage as a part of the i2010 strategy. In its 2006 recommendation the Commission recommended the creation of mechanisms to facilitate the use of orphan works [16]. Within the EDI efforts were made to give guidance to national measures to implement the recommendation. In 2007 a decision was made to develop sector-specific guidelines on due diligence criteria for orphan works. It was concluded that in addition to guidelines databases which contain information on orphan works would be beneficial. A first step in this respect was taken by the ARROW (Accessible Registries of Rights information and Orphan Works) project which gathers national libraries, collective management organizations and publishers and is co-funded by the European Commission under the eContent plus programme with €2.5 million. This project was launched in November 2008 to link different European rights registries as a first attempt.

One of the early statements on orphan works was made by the Libraries and Archives Copyright Alliance [17] in 2007. It gives a precise definition of an orphan work, explains the nature of an problem and offers several applicable solutions: licensing, diligent search, and mass digitization solution. The last one would allow not-for-profit educational, cultural and research institutions working with large numbers of orphan works to use them under a new exception. Issues to be considered with these solutions are discussed as a kind of practical guidance.

In October 2009 the European Commission adopted a Communication on Copyright in the Knowledge Economy aiming to tackle the important legal aspects of mass-scale digitization and dissemination of European libraries. The experiences gained with the Digital Library Europeana and consultations on the Green Paper were influential in the European Commission's putting challenges of book digitization for authors, libraries and consumers on the EU agenda [18].

5 User-Created Content

The Green Paper Copyright in the Knowledge Economy has also opened up a debate on knowledge dissemination in the online environment focusing on the relevant exceptions to copyright, which should include a possible exception for user-created content. Digital portals, considered to be particularly efficient places of knowledge dissemination, are expected to contain user-created content such as blogs, podcasts, wiki or video sharing, enabling users to easily create and share text, videos and pictures. Registered Europeana users, for instance, can save a search, add a tag, save an item or share it with a friend. Since the Directive [10] does not contain an exception which would allow the use of existing copyright protected content for creating new or derivative works, the obligation to clear rights before any transformative user-created content is made available is seen as an obstacle to the free dissemination of knowledge at present.

6 Digitization of Cultural Heritage in Croatia

Sporadic digitization of some valuable items or items of special interest in public institutions such as libraries, archives and museums took place as early as the late nineties of the 20th century. In 2006 an encompassing national programme of digitization of the cultural heritage kept in libraries, archives and museums throughout the country was initiated by the Ministry of Culture [19]. The first step was to gather all collections already digitized by cultural institutions in the country and display them on an electronic portal, allowing new collections to be added. The portal Croatian Cultural Heritage was developed as a central networking point that would enable accessibility and browsing of digital collections and as an efficacious means of coordination, i.e. the visibility of projects on the portal decreased the possible duplication of the material in the collections. In line with international documents, standards and guidelines for the digitization of material have been elaborated and institutional capacities for digitization strengthened. The portal has succeeded in gathering around 60 cultural and educational public institutions as participants. Since 2007 the number of collections displayed on the portal (<http://www.kultura.hr>) has been on the constant increase. The Ministry takes care of the portal and if needed offers technical support and advice. It also acts as a coordinator of the whole project. The institutions involved in the project have the freedom to select the material, prepare it for digitization, i.e. check completeness, do repairs if needed, add metadata, provide names of datafiles, etc. The technical part of the process is done either by professional firms or by departments of the institutions themselves, if they have at their disposal the adequate equipment. In 2010 a total of 34 digitization programs proposed by various public institutions have been approved and financed by the Ministry of Culture (a total of €720,000).

7 The Investigation

The aim of the investigation was to find out how copyright issues related to the digitization performed by public institutions such as libraries and archives have been dealt with in Croatia. This seemed to be interesting because little about that specific topic could be found on the portal itself and/or the cultural institutions' websites. Also, the national Copyright Act (2003) modelled in accordance with the EU Directive from 2001 does not explicitly mention digitization and is very cautious in providing exceptions to copyright protection. We were also aware that librarians and archivists in charge of digitization would have very little prior experience with copyright and we were curious to learn about their approach to the process of clearing rights.

The national portal *Croatian cultural heritage* was naturally our first source of information on how copyright is dealt with. But the portal simply claims that the rights to all the contents displayed belong to the Ministry of Culture. Moreover, in the same statement the Ministry waives any responsibility for the incorrectness or incompleteness of data. We found the statement to be somewhat ambiguous and made a decision to interview persons responsible for digitization projects in the four largest institutions

participating in the project, hoping that they might provide us with more inside knowledge. The institutions were: *the National and University Library in Zagreb*, the central library in the country, *Zagreb City Libraries*, the most important public libraries network in the country, the *Croatian State Archives*, and the *Croatian Academy of Sciences and Arts*. The fact that all institutions selected for the interviews are located in the capital might perhaps seem constraining, but at present they seem to be the leaders in digitization projects in the country. The interviews with persons responsible for digitization projects in those institutions were recorded on tape. The direct communication with colleagues helped us obtain answers that could not have been obtained from the portal and to get a better insight into the copyright situation. Thus the interview was not used as a method to obtain personal opinions, but as a means to clarify the ambiguous issues to which there was no explicit answer in other sources.

The number of digitization projects carried out by libraries, archives and museums in Croatia and displayed on the *Croatian Cultural Heritage* portal prove that Croatian librarians, archivists and museum curators follow the trends in other European countries trying to preserve and make available to the larger public the cultural heritage in their custody. However, considering the current hesitation of the EU public institutions to digitize 20th century material because of the expected obstacles in identifying and locating rights holders (see for instance [20]), one would expect the digitized 20th century material in Croatia to be scarce, too. However, even a short glance at the portal shows clearly that the majority of collections digitized (136) have been classified as containing 20th century materials. However, a closer inspection will also prove that those collections contain materials that span centuries and that only a part of the material in the collection belongs to the 20th century. Also, most of the 20th century collections seem to be collections of photographs, plans, posters and similar documents. But some collections house newspaper and journals clippings and some contemporary newspaper runs have also been digitized. It cannot be deduced from the portal itself how the rights have been cleared for every collection, apart from the already mentioned general copyright notice stating that the Ministry holds all rights for the materials displayed on the portal.

The colleagues in charge of digitization admit that the 20th century material has rarely been selected for digitization not only because of a fear that copyright clearance would require too much time and effort, but also because remuneration to rights holders would increase the envisaged expenses. Those answers led us to conclude that copyright remuneration was not foreseen as an expense in the project. The conclusion was corroborated by colleagues who confirmed that their institutions' project proposals did not contain any provision for copyright clearance expenses. However, in spite of the decision to avoid 20th century material, the Zagreb City Library came across a right that had to be cleared. The decision to digitize the first edition of a well-known classic children's book by an author whose rights expired a few years ago, proved to be unwise, because the edition also contained illustrations made by a painter who had died 20 years earlier and whose rights were still held by her niece. Although in the end the Library did not have to pay any royalties and the rights holder's only request was to obtain a digital copy, the length of the search for the right holder and the time and skill required to reach an agreement with her, proved to be a cumbersome task for the Library. Since there was no information on the illustrator in the library catalogue,

the search began with the Museum where the painter's retrospective exhibition was last organized almost 15 years ago, and the Museum's curators were indeed able to identify the present rights holder, but the search and further negotiations with the rights holder took a whole month. Taking into account the fact that externally funded digitization projects are to be carried out according to a predetermined time schedule, such inadvertent prolongation can be a serious impediment.

The Croatian Academy of Sciences and Arts, which has been a major publisher of scientific and research works in the country for the last hundred years decided to digitize its own publishing output first. It opted for open access, and the authors sign a contract and transfer their rights to the Academy. However, an option has been provided for the authors of older but still protected works to withhold the rights if they disagree with the policy of open access. This has not been the case, so far [21].

Interviews with colleagues revealed a sharp difference in the approach to digitization. For the librarians of the National and University Library digitization is primarily seen as a means of preservation; a new digital copy of a work is just another copy of the original held by the Library, although in a different format. In the words of the colleague the digital copy is made to preserve the original. On the contrary librarians in the Zagreb City Library strive to make an ideal copy; they try to 'embellish' the original not only by cleaning it from smudges, rust, etc., but also by removing all ex-libris, personal signatures and notes written on margins. If the Library copy is incomplete, they borrow a copy from another institution and amend their own copy. The Library's intention is to produce a new digital edition of the work. They also add a new imprint in the colophon of the digital copy and a new ISBN, and by doing that they openly claim that they are publishers of a new edition. Thus the City Library makes a new catalogue record, for a digital work, while the National and University Library just adds a note on the catalogue record of the original. Consequently, according to the Copyright Act the City Library could claim the rights on the new edition and perhaps even require remuneration for its use.

For the Croatian Academy of Sciences and Arts digitization is to a certain degree a means of preservation, but mostly it is an opportunity to make the works of its members accessible to a wider public. It allows the Academy to fulfill its main tasks: publish the results of scientific research and promote science.

In 2005 the National and University Library in Zagreb launched a website, Digitized Heritage, making a selection from the older national material as a part of the digitization project of the Library. Up to now more than 700 items have been digitized, mostly books, manuscripts and printed music, but also a collection of posters, drawings and maps as well as a few sound recordings. In spite of its initial decision to digitize only the material of national significance free of copyright, according to the librarian in charge of digitization, copyrighted material could not be completely avoided, and agreements have been concluded with artists (authors of posters) and performers of music on sound recordings. When permission is asked for, authors are approached individually, often by e-mail, and the authorization is in fact an e-mail reply received from the author. There is no special agreement concluded with the authors. The rights clearance for the sound recordings was obtained in writing by a collecting society responsible for music composers and performers.

A colleague librarian in charge of phonograms in the City Library discovered that she had to contact four different producers and the digitization of old gramophone records envisioned in 2008 had to be postponed to 2009.

The National and University Library also plans to digitize the 100 dissertations defended at the Zagreb University from the second half of the 19th century onwards. Some of those dissertations could still be protected, and the copyright holders should be found. The catalogue authority file which is normally consulted when basic data on authors are searched for is likely to be of no use in this case, because old catalogue records do not contain data on the authors' birth/death. However, the Lexicographic Institute in Zagreb, the publisher of numerous encyclopedias and lexicons, holds a database of biographical data on persons who had been active in various segments of social life in Croatia, researchers and scientists included, and the Institute is willing to make its database available to the National and University Library. But the database may be of help only in part, because what librarians really need are data on the transfer of rights and names and addresses of right holders and there is no such database in the country. The Croatian Society of Composers runs an agency, HDS/ZAMP, which is in fact a collecting society for music rights. But the data they hold are not available on the web and librarians have to put separate inquiries regarding an individual's rights to the HDS/ZAMP in writing.

The absence of any notice on orphan works or the possibility that an author whose work has been digitized may come forward after the digitization was done asserting his/her rights, has been conspicuous and proves that our librarians are still not aware of the full scope of copyright issues. Also, there is no information on the libraries' websites about possible false rights claims.

The City Library wanted to digitize a National Theatre Yearbook and Theatre Almanac and sent a request for permission to the National Theatre; they had to wait for an answer and succeeded in obtaining the permission after 7 months. This is interesting because it shows that the National Theatre as an important public institution has no plans to digitize its publications and/or archive for some time to come.

At its website the National and University Library has mounted a copyright notice stating that digital copies should not be published on the Internet or further distributed. They can be used for private purposes and research. Requests for commercial use should be sent in writing to the Library. Also, the users are obliged to cite the Library's portal as the source when using the material.

A great concern of colleagues in the National and University Library has been the fact that up to the present only material from special collections (maps, posters, music, old and rare books and manuscripts) has been selected for digitization. The general collection which includes the national collection *Croatica*, is digitized on demand only, meaning that if a user places an order for a digital copy of a work, the Library keeps a copy of its own. But this obviously means that the selection of the material to be digitized has been determined by the preferences of users.

The City Library provides a statement on the use of digital material. It allows the material to be used, but not for commercial purposes; downloaded material bears a watermark and can be used for reading but not for further reproducing.

Ambiguous copyright claims have also been noted. The National and University Library has paid royalties for digitization to the author of a catalogue of dissertations

published by an institute almost forty years ago, although the catalogue had been composed as a part of the author's regular duties as an employee of the institute. This case could be imitated in the future and employees of all kinds could claim rights for the works they created as part of their job obligations. In case of works on hire the current Copyright Act stipulates that the work contract should contain a clause on who has the rights. In the second half of the 20th century such clauses were exceptions.

The Croatian State Archives seems to be more interested in personal data protection than in copyright, although colleagues admit that they keep copyrighted material also. The Archives has very good equipment which allows for digitization of its own material as well as the material of other institutions. Up to now 1,600,000 pages of birth/death registries have been digitized and are available to the public. The Archives has not digitized sound or film materials, and textual material only appears on the portal Croatian Cultural Heritage. It is somewhat strange, since the Croatian Film Archive is a division of the State Archives. The Archives digitized the first Croatian film about a well-known Croatian composer, but a copy can be obtained on DVD only. Since authenticity is crucial for the archival documents, the Archives issues a written confirmation of authenticity for a digital copy of a document, but the user has to demand it expressly.

When asked where and how they learned about copyright related to the digitization process, our colleagues replied that they attended the courses on digitization held in the Centre for Continuing Education of Librarians in Zagreb and used the copyright primer for librarians recently published in Zagreb [22].

Although user-created content (UCC) and its relation to copyright and/or licensing is an issue much discussed in the literature at present (see for instance [23], [24], [25]) we were not able to learn much about it in our investigation. Apart from the answers to frequently asked questions and e-mail contact availability the Croatian Cultural Heritage portal also provides an opportunity for users to add their own content, but up to now nobody has done so. There is no blog and the site does not allow for user feedback.

Recently, however, the National and University Library has opened a Facebook page allowing both users and non-users to learn about its activities and programmes and comment on them. Users can send their proposals concerning content to be digitized to the Library.

8 Conclusion

A proliferation of EU programmes and action plans on digitization proves that there is a political will to digitize and make available to the public the rich cultural heritage of Europe. However, there are still obstacles to overcome, such as the high cost, lack of technical standards and lack of consistency in approaches to intellectual property rights. Since 2000, Croatian public institutions such as libraries, archives and museums have started digitizing their collections, but little can be learned from public sources about the solutions taken regarding rights clearing, so an investigation has been carried out to find out more details and compare the situation with trends in the EU. The interviews held with persons responsible for digitization projects in four

major public institutions in Croatia proved that Croatian librarians and archivists are aware of the fact that digitization is closely connected to copyright issues and that rights have to be cleared. They are also aware that they must investigate the copyright situation and legal position affecting access by users to the material created by the project. They have been able to gain a basic knowledge of copyright issues related to digitization by attending the courses in the Centre for Continuous Education of Librarians and by using a recently published book on copyright for librarians written in Croatian. Library catalogues serve as the first source of information on authors, but they are helpful only in determining if a work is still protected and lack all other information needed to trace and locate authors. A lack of suitable registries or databases with data on authors in the country is a further serious obstacle, and the only exception is a database on composers and performers kept by a collecting society for music rights. Institutions show different approaches to their own role; the Croatian Academy of Sciences and Arts and Zagreb City Libraries consider themselves to be publishers and rights holders of new digital editions of works, while the National and University Library and the State Archive take digitization primarily as a means of protection of original works.

As a rule, institutions have a policy of copyright management, even if a rudimentary one. All institutions provide copyright information, sometimes with a copyright disclaimer. They use watermarking as a technical protection measure. Digitized materials are provided at no cost on the Internet for private use and research. Use of materials for commercial purposes has to be paid and the institutions look upon it as a source of revenue to fund further digitization activities. A definitely distinct policy has been adopted by the Academy of Sciences and Arts which publishes results of research of its own members and which has decided to join the open access initiative. The policy of all institutions toward derivative works, i.e. offering the opportunity to users to create their own content and add/mix their contributions with the content digitized by the institutions is yet to be determined.

References

1. Council and European Commission for the Feira Council: e-Europe2002 Action Plan (2000), http://ec.europa.eu/information_society/eeurope/2002/action_plan/pdf/actionplan_en.pdf
2. Lund principles: Lund Principles: Conclusion of Experts Meeting, Lund, Sweden (April 4, 2001), <http://cordis.europa.eu/ist/digicult/lund-principles.htm>
3. Lund Action Plan, ftp://ftp.cordis.europa.eu/pub/ist/docs/digicult/lund_action_plan-en.pdf
4. Charter of Parma, <http://www.minervaeurope.org/structure/nrg/documents/charterparma031119final.htm>
5. Dynamic action plan for the EU co-ordination of digitization of cultural and scientific content, <http://www.minervaeurope.org/publications/dap/dap.pdf>
6. Ricolfi, M., et al.: Final Report on Digital Preservation, Orphan Works, and Out-of Print Works (i2010: Digital libraries High Level Expert Group – Copyright Subgroup) (2008), http://ec.europa.eu/information_society/activities/

- digital_libraries/doc/hleg/reports/copyright/
copyright_subgroup_final_report_26508-clean171.pdf
7. Cultural Heritage, Information Society Technologies,
<http://cordis.europa.eu/ist/digicult/eeurope.htm>
 8. Berne Convention for the Protection of Literary and Artistic Works,
<http://www.wipo.int/treaties/en/ip/berne/>
 9. Copyright in the Knowledge Economy: Green Paper. Brussels: Commission of the European Communities, (COM (2008)466/3),
[http://ec.europa.eu/internal_market/copyright/docs/
copyright-infso/greenpaper_en.pdf](http://ec.europa.eu/internal_market/copyright/docs/copyright-infso/greenpaper_en.pdf)
 10. Directive 2001/29/EC of the European Parliament and of the Council of May 22, 2001 on the Harmonisation of Certain Aspects of Copyright and Related Rights in the Information Society, <http://www.ivir.nl/legislation/eu/copyright-directive.doc>
 11. IFLA: IFLA Guidelines for Digitization Projects for Collections and Holdings in the Public Domain, Particularly those Held by Libraries and Archives (2002),
<http://www.ifla.org>
 12. EIFL, IFLA and LCA Statement of Principles on Copyright Exceptions and Limitations for Libraries and Archives,
[http://www.eifl.net/cps/sections/docs/ip_docs/statement-
principles/downloadFile/attachedFile_f0/
StatementofPrinciples_English.pdf?nocache=1261508729.81](http://www.eifl.net/cps/sections/docs/ip_docs/statement-principles/downloadFile/attachedFile_f0/StatementofPrinciples_English.pdf?nocache=1261508729.81)
 13. WIPO SCCR: Analytical document on limitations and exceptions. WIPO Standing Committee on Copyright and Related Rights, Nineteenth session, Geneva, SCCR19/3 (December 14 - 18, 2009),
http://www.wipo.int/meetings/en/doc_details.jsp?doc_id=129524
 14. Bide, M., Oppenheim, C., Ramsden, A.: Some Proposals Regarding Copyright Clearance and Digitisation in Higher Education. *Journal of Information Science* 23, 393-405 (1997)
 15. In from the Cold: An Assessment of the Scope of 'Orphan Works' and its Impact on the Delivery of Services to the Public,
<http://www.jisc.ac.uk/contentalliance/>
 16. Commission of the European Communities: Commission Recommendation of August 24, 2006 on the Digitisation and Online Accessibility of Cultural Material and Digital Preservation, Brussels, [http://eur-lex.europa.eu/LexUriServ/
LexUriServ.do?uri=OJ:L:2006:236:0028:0030:EN:PDF](http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:236:0028:0030:EN:PDF)
 17. LACA: Statement on Orphan Works,
[http://www.cilip.org.uk/sitecollectiondocuments/PDFs/policya
dvocacy/laca/LACAorphanworksstatementFINAL19dec07.pdf](http://www.cilip.org.uk/sitecollectiondocuments/PDFs/policyadvocacy/laca/LACAorphanworksstatementFINAL19dec07.pdf)
 18. Communication from the Commission: Communication on Copyright in the Knowledge Economy. Commission of the European Communities, (COM(2009)532final), Brussels (2009), [http://ec.europa.eu/internal_market/copyright/docs/
copyright-infso/20091019_532_en.pdf](http://ec.europa.eu/internal_market/copyright/docs/copyright-infso/20091019_532_en.pdf)
 19. Ministry of Culture: National Programme of Digitisation of Archival, Library and Museum Holdings: Summary. Zagreb (2007),
[http://www.kultura.hr/eng/Join-in/Download-documents/
The-National-Programme-of-Digitisation-of-Archival-Library-
and-Museum-Holdings](http://www.kultura.hr/eng/Join-in/Download-documents/The-National-Programme-of-Digitisation-of-Archival-Library-and-Museum-Holdings)
 20. LIBER Response to the EC Creative Content Online Consultation,
[http://www.libereurope.eu/files/
LIBER-Response-Creative-Content-Online-Consultation.pdf](http://www.libereurope.eu/files/LIBER-Response-Creative-Content-Online-Consultation.pdf)

21. Crnkovic, K., Juricic, V., Polak Bobic, K.: Digitalna Zbirka Knjiznice Hrvatske Akademije Znanosti i Umjetnost. Croatian Cultural Heritage, <http://www.kultura.hr>
22. Horvat, A., Zivkovic, D.: Knjiznice I Autorsko Pravo (Libraries and Copyright). Hrvatska sveucilisna naklada, Zagreb (2009)
23. Helberger, N., et al.: Legal Aspects of UCC (2010), <http://www.ivir.nl/index.html>
24. Halbert, D.: Mass Culture and the Culture of the Masses: A Manifesto for User-Generated Rights. *Vanderbilt Journal of Entertainment & Technology Law* 11, 921–961 (2009)
25. Gervais, D.: The Tangled Web of UGC: Making Copyright Sense of User-Generated Content. *Vanderbilt Journal of Entertainment & Technology Law* 11, 841–870 (2009)

Performance Evaluation of Web Information Systems

Yaşar Tonta and İrem Soydal

Hacettepe University, Department of Information Management, 06800, Ankara, Turkey
{tonta, soydal}@hacettepe.edu.tr

Abstract. The quality of web information systems is closely related with their users' expectations. In order to create good quality systems, web designers should understand how users perceive service quality of various performance attributes such as security, usability and information quality, and which ones they value the most. In this study, users' service quality perceptions of two different web sites were investigated. Data were obtained from 1,900 users of web sites of a for-profit online bookstore and a not-for-profit national information center through the E-Qual Index that was administered online. Users were asked to rate to what extent each web site satisfied the 22 performance attributes included in the E-Qual Index along with the importance they attached thereto. Quadrant analysis and gap analysis were used to analyze the data to find out the strengths and weaknesses of the web sites investigated. Moreover, performance attributes that need to be improved or that were "undervalued" by the users were identified. Findings indicate that each web site satisfied more than half the performance attributes listed in the E-Qual Index. Users attached more importance to performance attributes such as ease of use, provision of timely and relevant information and safety. Findings can be used to remedy the shortcomings of each web site and improve the web service quality.

Keywords: Web service quality, E-Qual index, quadrant analysis, gap analysis, usability, information quality.

1 Introduction

Web sites presenting information about a specific subject that appeal to a specific user group are called web information systems. These systems may have commercial, information sharing or information presenting purposes, and they need to be designed taking into consideration user satisfaction. The level of user satisfaction is determined by the extent to which users' expectations are met. Users often think that the quality of web information systems is closely related with the quality of information provided and the quality of system design [1], [2], [3]. Users' expectations increase as web functionality does. Users become more demanding as they experience new web services and technologies [4]. This shows that users' expectations can change as fast as the Web itself does. Measuring service quality in the Web environment is important in terms of increasing the usage rate of systems. If web service quality is perceived positively by

the users, it significantly affects the users' overall satisfaction, their eagerness to suggest the system to others and their tendencies to purchase/repurchase [5].

This study attempts to measure the perceived service quality of web information systems by means of the E-Qual Index. It aims to understand users' differing approaches to web service quality attributes by comparing two web sites as examples. Comprised of 22 Likert-type questions, the E-Qual Index tests the concepts of web site usability, information quality and interaction quality, in general. The E-Qual Index was used to evaluate the users' perceived service quality of one for-profit web site (Idefix, an online shopping site, www.idefix.com.tr) and one not-for-profit web site (the National Academic Network and Information Center, ULAKBIM, www.ulakbim.gov.tr). Findings were compared to see if users' perceptions of web service quality differed in terms of types of web sites. Strengths and weaknesses of each web site were also identified.

2 Literature Review

The concept of "traditional" service quality started to attract attention in the 1980s [6]. The SERVQUAL Index with 22 questions and five different dimensions (Tangibles, Reliability, Responsiveness, Assurance, and Empathy) emerged as a tool to measure the traditional service quality. The dimensions measured by the SERVQUAL Index have since changed as more services and products have increasingly been offered via the Web. Still, most studies measured web service quality by using dimensions of the SERVQUAL Index [3], [7], [8], [9].

Several studies have been carried out on web service quality [2], [10], [11], [12], [13]. Factor analysis was used to identify the new web service quality dimensions. Among them are: usability, usefulness of content, the adequacy of information, accessibility, and interaction [12]; responsiveness, competence, quality of information, empathy, web assistance, and callback systems [14]; and, effectiveness, system availability, fulfillment, compensation, and contact [15]. Also, the quality of information, the features of perceived usability and the ease of use are significant measures from the users' point of view to evaluate the success of information systems [1], [2].

The E-Qual Index was developed at the beginning of the 2000s in order to measure web service quality. It was based on the literature of information systems, marketing and human-computer interaction [16]. First versions of E-Qual Index (named "Web-Qual Index" before 2003) were used to evaluate web sites of various universities [17], online auction web sites [18], web sites of online bookstores [16], government institutions providing electronic services [19], [20] and an online bookstore and a national information center [21], [22]. These studies show that E-Qual Index is a kind of "built-in scale" and reveals a consistent methodology and some harmonious results. Using factor analysis on data, dimensions of perceived service quality were usually identified as Usability, Design, Information Quality, Trust and Empathy. In one of these studies, the information presenting website of OECD's Forum on Strategic Management Knowledge Exchange (FSMK) was explored by removing the questions on Trust [23]. Results of the study were similar except for the Trust dimension. Users' views on Usability and Design dimensions changed positively after the web site was

redesigned. A similar study was carried out using the official website of the UK Inland Revenue Service in which users' perceptions of service quality were compared. Users who utilized the web site to carry out "interactive online transactions" attached more importance to the Usability dimension than those who used it to "gather information" [19], [20]. Another study investigated the differences of perceived service quality of a for-profit and a not-for-profit web site [21], [22]. Factor analysis results showed that the dimensions Usability, Design and Empathy came up almost the same for both web sites. However, the not-for-profit web site users perceived the "Trust" issues merged with the "Information Quality" and thought that this consolidated dimension was the most important service quality component. On the other hand, the for-profit web site users comprehend Trust and Information Quality as different dimensions and thought the Trust dimension more important.

3 Methodology

Research questions addressed in this study are as follows:

- How do users perceive the service quality of for-profit and not-for-profit web sites in terms of various performance attributes such as information quality, usability and service interaction quality?
- Which performance attributes do users value most as part of the web service quality?
- What are the strengths and weaknesses of for-profit and not-for-profit web sites in terms of design?
- Are there any performance attributes that can be improved?
- Are there any performance attributes that are "undervalued" by the users?

The Web sites of Idefix, a for-profit online bookstore, and ULAKBIM, a not-for-profit national information center, were used to collect data using the E-Qual Index (www.webqual.co.uk/instrument.htm). An online questionnaire with 22 questions was sent to all the registered users of both web sites via e-mail. Questions addressed the aspects of usability (e.g., "I find the website easy to learn to operate"), information quality (e.g., "The web site provides accurate/timely/believable information") and service interaction (e.g., "My personal information feels secure") of each web site. Users were asked to rate to what extent they were pleased with the service quality delivered by each web site for each performance attribute (e.g., their perceived satisfaction level) and how much importance they would attach thereto. They marked their ratings of satisfaction and importance for each question on a 5-point Likert scale (1: "I'm not pleased at all" / "It is not important for me at all" - 5: "I'm very pleased" / "It is very important for me"). The perceived satisfaction scores reveal the users' evaluation of each web site while the importance scores reveal their levels of expectations for performance attributes tested in terms of web service quality.

Analyses were based on 1,782 and 118 questionnaires filled out by Idefix and ULAKBIM web site users, respectively. Findings were summarized in tables and quadrant graphs.

Quadrant analysis ... is a graphic correlation technique that produces data easy to visualize ... The technique plots data about service attributes into four quadrants defined by two dimensions: one reflects the importance to service excellence that customers give service attributes, while the other indicates the extent to which customers think a particular service has the attributes. The first dimension is plotted along the horizontal axis as the ideal expectation for excellent service quality, and the second dimension is plotted along the vertical axis as the perceived ... performance. [24]

This type of importance-performance analysis lets the system designers or managers identify their action plans to increase the overall satisfaction level of systems' users [25].

In this study, quadrant graphs were used to plot the 22 performance attributes of each web site by users' ratings of importance and satisfaction. Mean scores of importance and satisfaction for each attribute were plotted on a quadrant graph.

A quadrant graph has four partitions, each with its own specific interpretation. The upper right quadrant (Quadrant 1) contains performance attributes of "high importance-high satisfaction" levels representing the strengths of the web site (so called "winners"). These attributes need to be retained by system designers.

The upper left quadrant (Quadrant 2) represents the "low importance-high satisfaction" performance attributes, so called "undervalued" ones [26]. System designers should rethink the resources allocated to these services or they should figure out how they can improve the images of these performance attributes so that users think that they are important (like the ones in Quadrant 1) [24].

The lower left quadrant (Quadrant 3) denotes the "low importance-low satisfaction" performance attributes. These are relatively less important performance attributes in the eyes of users with relatively lower satisfaction levels attained (i.e., "unimportant weaknesses" that can be ignored). The investment that has been made in them should be re-examined and resources perhaps be transferred to improve other performance attributes that are perceived as more important by the users [24].

Performance attributes with "high importance-low satisfaction" levels are located in the lower right quadrant (Quadrant 4). The attributes falling into Quadrant 4 are also very important for users but they do not think the system has an outstanding performance in terms of these attributes. Performance attributes in Quadrant 4 ("Opportunity quadrant") are candidates for immediate attention of web site designers to improve service quality so that users' expectations can be met more successfully [26].

Gaps that exist between importance and satisfaction were identified by gap analysis. "Large gaps between importance and satisfaction rating usually draw attention to problems that must be corrected. Small gaps signify strengths." [26].

We present the findings of our study on a gap analysis graph and quadrant charts to identify strengths and weaknesses of web sites as well as to find out the "undervalued" or "overdone" performance attributes. Based on the outcome of quadrant analysis and gap analysis, it is possible to identify performance attributes that should be retained, improved, revisited or simply ignored.

4 Findings and Discussion

Mean importance and satisfaction scores for each performance attribute for both the Idefix and ULAKBIM web sites are given in Table 1.

Table 1. Mean importance and satisfaction scores

Attributes	Idefix		ULAKBIM	
	Imp.	Sat.	Imp.	Sat.
1. I find the site easy to learn to operate	4.4	4.3	4.5	3.7
2. My interaction with the site is clear and understandable	4.3	4.3	4.4	3.6
3. I find the site easy to navigate	4.6	4.1	4.6	3.5
4. I find the site easy to use	4.6	4.1	4.6	3.6
5. The site has an attractive appearance	3.8	3.3	3.7	2.8
6. The design is appropriate to the type of site	4.0	3.7	3.9	3.3
7. The site conveys a sense of competency	4.0	3.7	4.0	3.4
8. The site creates a positive experience for me	3.8	3.8	4.1	3.5
9. Provides accurate information	4.8	4.2	4.8	4.3
10. Provides believable information	4.7	4.1	4.8	4.3
11. Provides timely information	4.7	4.1	4.8	4.0
12. Provides relevant information	4.6	4.0	4.8	3.9
13. Provides easy to understand information	4.6	4.2	4.7	3.8
14. Provides information at the right level of detail	4.6	3.7	4.6	3.8
15. Presents the information in an appropriate format	4.5	4.0	4.6	3.8
16. Has a good reputation	4.5	4.3	4.5	4.2
17. It feels safe to complete transactions	4.8	4.4	4.6	4.1
18. My personal information feels secure	4.9	4.2	4.7	4.1
19. Creates a sense of personalization	3.9	3.3	3.9	3.2
20. Conveys a sense of community	3.0	2.9	3.7	3.5
21. Makes it easy to communicate with the organization	4.2	3.7	4.3	3.8
22. I feel confident that goods/services will be delivered as promised	4.8	4.2	4.6	4.1
\bar{X} :	4.4	3.9	4.4	3.7

Note: "Imp." and "Sat." stand for mean importance and mean satisfaction scores, respectively. \bar{X} represents the average of mean scores.

On a five-point scale ranging from 1 ("not important at all") to 5 ("very important"), performance attributes that received the highest mean scores of importance from Idefix users were as follows (the average of means was 4.4):

"18. My personal information feels secure" (mean 4.9);

"9. Provides accurate information" (mean 4.8);

"17. It feels safe to complete transactions" (mean 4.8);

"22. I feel confident that goods/services will be delivered as promised" (mean 4.8);

"10. [The web site] Provides believable information" (mean 4.7);

"11. [The web site] Provides timely information" (mean 4.7).

Performance attributes that received the highest mean scores of importance from ULAKBIM users were as follows (the average of means was 4.4):

- “9. [The web site] Provides accurate information” (mean 4.8);
- “10. [The web site] Provides believable information” (mean 4.8);
- “11. [The web site] Provides timely information” (mean 4.8);
- “12. [The web site] Provides relevant information” (mean 4.8);
- “13. [The web site] Provides easy to understand information” (mean 4.7);
- “18. My personal information feels secure” (mean 4.7).

It is clear that web site users of the for-profit online bookstore (Idefix) attached tremendous importance to security, safety and confidence issues as well as to receiving believable and timely information. Receiving accurate, believable, timely, relevant and easy to understand information were also the most important performance attributes for web site users of the not-for-profit information center (ULAKBIM). They were also concerned with the security of their personal information held by the information center.

Statements that received the lowest mean scores of importance from Idefix and ULAKBIM users were very similar to each other, although the order of importance for ULAKBIM was slightly different (the order of number 6 and 7 was reversed and number 8 was not among the lowest mean scores of ULAKBIM):

- “6. The design is appropriate to the type of site” (Idefix mean: 4.0; ULAKBIM mean: 3.9);
- “7. The site conveys a sense of competency” (Idefix mean: 4.0; ULAKBIM mean: 4.0);
- “19. Creates a sense of personalization” (Idefix mean: 3.9; ULAKBIM mean: 3.9);
- “5. The site has an attractive appearance” (Idefix mean: 3.8; ULAKBIM mean: 3.7);
- “8. The site creates a positive experience for me” (Idefix mean: 3.8);
- “20. Conveys a sense of community” (Idefix mean: 3.0; ULAKBIM mean: 3.7).

Users of both the for-profit and not-for-profit web sites attached much less importance to such usability and interaction quality issues as attractive appearance of the web site or the availability of personalization features.

As indicated earlier, users were also asked to rate their perceived satisfaction levels for each performance attribute of the web site that they used (e.g., Idefix’s or ULAKBIM’s web site). The highest mean scores of perceived satisfaction for Idefix users were as follows (average of means was 3.9):

- “17. It feels safe to complete transactions” (mean 4.4);
- “1. I find the site easy to learn to operate” (mean 4.3);
- “2. My interaction with the site is clear and understandable” (mean 4.3);
- “16. [The web site] Has a good reputation” (mean 4.3);
- “9. [The web site] Provides accurate information” (mean 4.2);
- “13. [The web site] Provides easy to understand information” (mean 4.2);
- “18. My personal information feels secure” (mean 4.2);
- “22. I feel confident that goods/services will be delivered as promised” (mean 4.2).

Corresponding highest mean scores of perceived satisfaction for ULAKBIM users were as follows (average of means was 3.7):

- “9. [The web site] Provides accurate information” (mean 4.3);
- “10. [The web site] Provides believable information” (mean 4.3);
- “16. [The web site] Has a good reputation” (mean 4.2);
- “17. It feels safe to complete transactions” (mean 4.1);
- “18. My personal information feels secure” (mean 4.1);
- “22. I feel confident that goods/services will be delivered as promised” (mean 4.1).

Users of both web sites found systems' reputation, safety and security most satisfactory. They highly trusted the accuracy of information provided. ULAKBIM web site users perceived it very easy to communicate with ULAKBIM while Idefix users highly appreciated the ease of operation and their straightforward interaction with the web site.

Two out of three performance attributes that were rated as the least satisfactory by users of these web sites were in common (numbers represent Idefix's order of importance along with the lowest mean scores for both web sites):

- “5. The site has an attractive appearance” (Idefix mean: 3.3; ULAKBIM mean: 2.8);
- “19. [The web site] Creates a sense of personalization” (Idefix mean: 3.3; ULAKBIM mean: 3.2);
- “20. [The web site] Conveys a sense of community” (Idefix mean: 2.9; ULAKBIM mean: 3.5).

Users did not find the web sites attractive at all and they thought that the personalization features available on the web sites were quite unsatisfactory.

It appears that all performance attributes included in the E-Qual Index were important to users and the mean of means of importance for both Idefix ($\bar{X}=4.4$, $SD=0.8$) and ULAKBIM ($\bar{X}=4.4$, $SD=0.7$) were higher than those of satisfaction (Idefix: $\bar{X}=3.9$, $SD=1.0$; ULAKBIM: $\bar{X}=3.7$, $SD=1.2$).

No statistically significant correlation was observed between the ranks of importance and ranks of satisfaction scores of Idefix and ULAKBIM web site users. (Spearman's correlation coefficient was greater than .05 in each case.) That is to say, users of each web site attached importance to relatively different performance attributes. Similarly, there was no statistically significant correlation between the ranks of each web site's importance and satisfaction scores. Users were not necessarily satisfied with the performance attributes that they rated highly important and vice versa.

As both Idefix and ULAKBIM data were not normally distributed (p values for Kolmogorov-Smirnov tests were below .05), a nonparametric (Wilcoxon signed rank) test was applied. Differences between the mean scores of importance and satisfaction for both ULAKBIM and Idefix web sites users were statistically significant for all but three performance attributes (question numbers 8, 16 and 20). The following statements did not elicit statistically significant differences between importance and satisfaction mean scores:

- “20. [The web site] Conveys a sense of community” (Idefix and ULAKBIM);
- “8. The site creates a positive experience for me” (Idefix); and
- “16. [The web site] Has a good reputation for me” (ULAKBIM).

Minimum and maximum distances from the mean scores of importance and satisfaction for performance attributes seemed to vary for Idefix and ULAKBIM users. For instance, security (question no. 18) was considered the most important performance attribute by Idefix users (4.9 points out of 5.0, the average being 4.4) while accurate, believable, timely and relevant information (question nos. 9, 10, 11 and 12) were the most important attributes for ULAKBIM users (4.8 out of 5, average being 4.4). The web site conveying a sense of community (question no. 20) was considered the least important attribute for both Idefix and ULAKBIM users (3.0 and 3.7 points, respectively). The attractive appearance of a web site (question no. 5) was also considered equally least important by ULAKBIM users (3.7). Idefix users were highly satisfied (4.4, average being 3.9) with the safety of Idefix web site (question no. 17) whereas ULAKBIM users found accurate and believable information (question nos. 9 and 10) furnished by ULAKBIM web site quite satisfactory (4.3, average being 3.7). Idefix users found the Idefix web site’s conveyance of a sense of community (question no. 20) much less satisfactory (2.9) and ULAKBIM users reckoned ULAKBIM’s web site much less attractive (average 2.8, question no. 5).

4.1 Quadrant Analysis

Distances from the mean scores of importance and satisfaction for each performance attribute were plotted on a quadrant graph for the Idefix web site (Fig. 1).

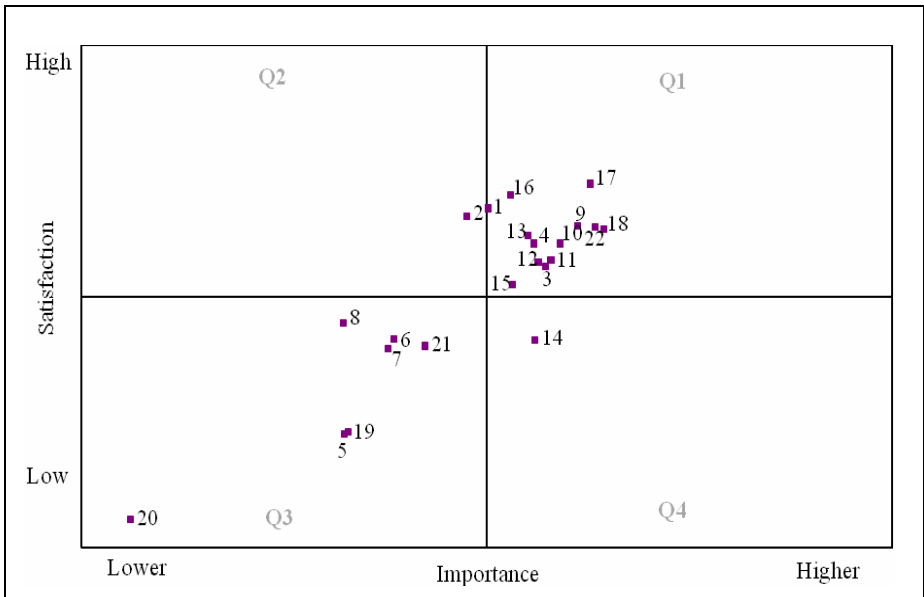


Fig. 1. Quadrant graph of performance evaluation of Idefix web site (■’s correspond to mean scores of the attribute numbers in Table 1)

Quadrant 1 (Q1) contains performance attributes which Idefix users rated highly important and that they were highly satisfied with. To put it differently, performance

attributes in Q1 denote the strengths of the Idefix web site. These attributes were as follows (numbers correspond to question numbers in Table 1):

- “1. I find the site easy to learn to operate”;
- “3. I find the site easy to navigate”;
- “4. I find the site easy to use”;
- “9. The site creates a positive experience for me”;
- “10. [The web site] Provides believable information”;
- “11. [The web site] Provides timely information”;
- “12. [The web site] Provides relevant information”;
- “13. [The web site] Provides easy to understand information”;
- “15. [The web site] Presents the information in an appropriate format”;
- “16. [The web site] Has a good reputation”;
- “17. It feels safe to complete transactions”;
- “18. My personal information feels secure”;
- “22. I feel confident that goods/services will be delivered as promised”.

As the list suggests, the Idefix web site has some strong performance attributes related with usability (nos. 1, 3-4), information quality (nos. 9-13, 15) and interaction quality (nos. 16-18, 22).

Quadrant 2 (Q2) represents “low importance-high satisfaction” performance attributes and the Idefix web site has only one performance attribute pertaining to usability in Q2 (no. 2: “My interaction with the site is clear and understandable”). It seems that users “undervalued” this performance attribute [26]. To put it differently, efforts expended by Idefix web site designers exceeded users’ expectations (i.e., “overkill”) [27].

Quadrant 3 (Q3) represents “low importance-low satisfaction” performance attributes of Idefix web site. These were as follows:

- “5. The site has an attractive appearance”;
- “6. The design is appropriate to the type of site”;
- “7. The site conveys a sense of competency”;
- “8. The site creates a positive experience for me”;
- “19. [The web site] Creates a sense of personalization”;
- “20. [The web site] Conveys a sense of community”;
- “21. [The web site] Makes it easy to communicate with the organization”.

Performance attributes in Q3 were seen as relatively unimportant by Idefix users. It is interesting to note that performance attributes of personalization, sense of community and easy communication with the organization (nos. 19-21) were not seen as very important by Idefix users but they were not terribly impressed by what the Idefix web site delivered in this respect.

Quadrant 4 (Q4) represents “high importance-low satisfaction” performance attributes. Idefix has one performance attribute related with information quality in Q4 (no. 14: “[The web site] Provides information at the right level of detail”). Idefix users did not seem to be satisfied with the level of detail that Idefix web site provided. Designers can benefit from users’ feedback to improve the level of satisfaction for this performance attribute.

Fig. 2 provides the quadrant graph for the ULAKBIM web site.

Performance attributes of the ULAKBIM web site contained in Q1 are similar to those of the Idefix web site (nos. 1, 9-13, 15-18, 22) with the exceptions that the ULAKBIM graph (a) lacks performance attributes pertaining to usability such as ease of navigation and ease of use (nos. 3-4) in Q1; and (b) has an additional performance attribute of information quality (no. 14).

Q2 has one performance attribute (no. 21: “[The web site] Makes it easy to communicate with the organization”) that represents low importance but high satisfaction.

ULAKBIM’s efforts to improve performance for this attribute seem to have been “undervalued” by its users.

Q3 has six performance attributes representing both usability (nos. 5-8) and service interaction (nos. 19-20):

- “5. The site has an attractive appearance”;
- “6. The design is appropriate to the type of site”;
- “7. The site conveys a sense of competency”;
- “8. [The web site] Creates a positive experience for me”;
- “19. [The web site] Creates a sense of personalization”;
- “20. [The web site] Conveys a sense of community”.

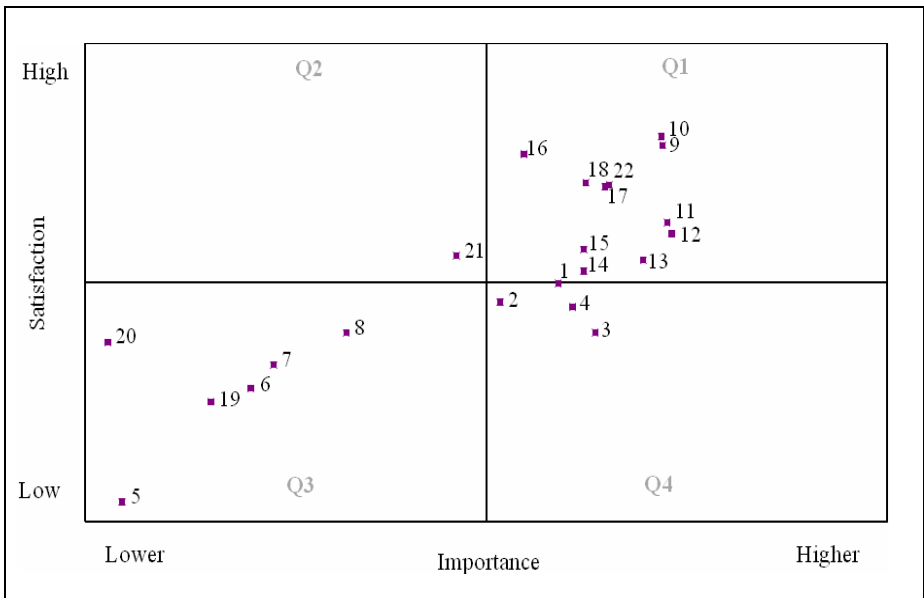


Fig. 2. Quadrant graph of performance evaluation of ULAKBIM web site (■’s correspond to mean scores of the attribute numbers in Table 1)

ULAKBIM users seem not to have minded the weaknesses of these attributes.

Q4 has three performance attributes related with usability (nos. 2-4) of ULAKBIM web sites:

- “2. My interaction with the site is clear and understandable”;
- “3. I find the site easy to navigate”;
- “4. I find the site easy to use”.

In users' eyes, these attributes carry high importance but users did not think it was easy to navigate or use the web site. Thus, the immediate attention of ULAKBIM's web site designers is needed.

Results of quadrant analysis suggest that both the Idefix and the ULAKBIM web sites satisfy more than half the performance attributes listed in the E-Qual Index. Both Idefix and ULAKBIM users seem to attach higher importance to attributes such as ease of use, provision of timely and relevant information, and safety. Performance attributes listed in Q1 are the strengths of both web sites.

Clear and understandable interaction that the Idefix web site offers to its users, and the ease with which users can communicate with ULAKBIM through its web site seem not to have been appreciated enough by web site users, as they undervalued these two performance attributes (Q2). It could be that these two attributes of interaction quality and usability are considered as a "given" by users and they tend to expect such performance criteria to be met by all web sites.

It is interesting to note that both Idefix and ULAKBIM users seem to attach less importance to such attributes as attractive appearance of web site, competency, positive experience, personalization, and sense of community. Both web sites delivered a service that was, in their users' eyes, somewhat below the average for performance attributes in Q3. In other words, users found these weaknesses not terribly important.

As indicated earlier, Q4 contains high importance but low satisfaction performance attributes. Idefix users did not think the web site provided information at the right level of detail while ULAKBIM users did not find the web site easy to navigate and use. Users' feedback provides an opportunity for Idefix and ULAKBIM web site designers to correct these shortcomings and be appreciated by their users.

4.2 Gap Analysis

Quadrant analysis "does not explicitly identify gaps that may exist between importance and satisfaction". "Even though a performance attribute appeared in the "high importance and high satisfaction" quadrant, a large gap could exist between importance and satisfaction ratings." [27]

A gap analysis was carried out to identify large gaps (i.e., shortcomings to be fixed), small gaps (i.e., strengths) and negative gaps (i.e., "overkill"). As indicated earlier, users, in general, attached consistently higher importance to performance attributes than their perceived satisfaction levels. Fig. 3 illustrates the gaps between importance and satisfaction scores for both Idefix and ULAKBIM web site users.

Gaps between importance and satisfaction scores were, in general, larger for the ULAKBIM web site ($\bar{X}=0.7$, $SD=0.2$) than those of the Idefix web site ($\bar{X}=0.4$, $SD=0.2$), indicating that expectations of Idefix users were met more successfully. The largest gaps between importance and satisfaction scores for the ULAKBIM web site (and therefore the most problematic performance attributes) were as follows: easy navigation, usable, understandable information, attractive and relevant information (questions 3, 4, 13, 5 and 12). The most problematic performance attributes of the Idefix web site were the following: right detail of information, security, keeps promises, timely and relevant information (questions 14, 18, 22, 11 and 12). Performance attributes with the smallest gaps (and thus representing the strengths) were reputation, sense of community and believable information for the ULAKBIM web site (questions

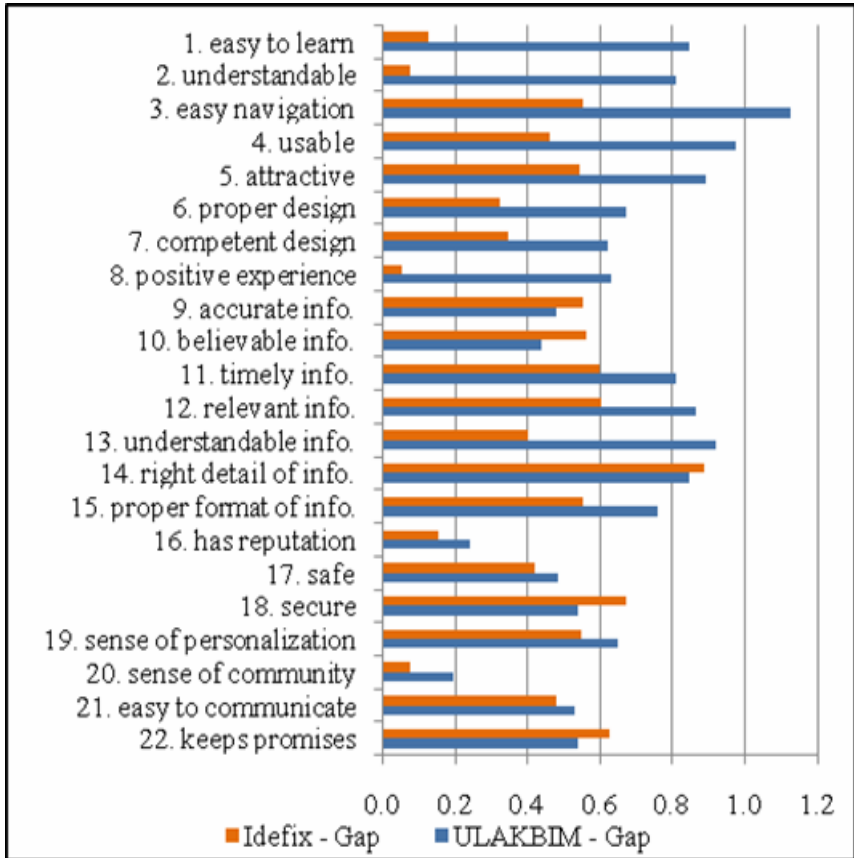


Fig. 3. Gap analysis for Idefix and ULAKBIM

16, 20 and 10), and positive experience, sense of community, understandable, easy to learn and reputation for the Idefix web site (questions 8, 20, 2, 1 and 16). Note that no negative gaps were observed for either web site, indicating that users’ average satisfaction scores were never higher than those of importance scores for any performance attribute.

Results of gap analysis suggest that the ULAKBIM web site has navigation and usability problems, is not attractive and does not offer understandable and relevant information according to its users. The strengths of the web site are that it has reputation, creates a sense of community and offers believable information. The Idefix web site, on the other hand, does not offer timely and relevant information with the right detail, has problems with keeping its promises of delivery of goods and security. Yet, users are satisfied with the positive experience that the Idefix web site offers along with its reputation, understandable and easy to learn user interface.

5 Conclusion

This study was carried out to find out users' perceptions of service quality for two different web sites, one for-profit (Idefix) and the other not-for-profit (ULAKBIM). It also aimed to investigate the perceived levels of satisfaction of service quality attributes and their importance as rated by the users.

Findings indicate that users' expectations were not fully met by either web site. Yet users were satisfied with service quality of more than half the performance attributes such as ease of use, provision of timely and relevant information, and safety and security, to which they also attached the greatest importance. For-profit web site users were more concerned about safety and security of the systems they used while not-for-profit users valued accurate, timely and relevant information. Users were much less concerned with attractive appearance of web sites regardless of the type of web site (i.e., for-profit or not-for profit). They were most satisfied with service quality of safety and security offered by web sites but highly dissatisfied with personalization features available (or lack thereof).

ULAKBIM web site users did not find the web site easy to navigate and use, although they thought these performance attributes quite important. Similarly, Idefix users were not pleased with the level of detail of information presented by the web site. These performance attributes should be revisited to meet the service quality expected by the users.

Some performance attributes such as clear and understandable interaction and easy communication seem to have been undervalued by users. They were more forgiving of, and attached relatively less importance to, service quality provided by web sites for performance attributes such as personalization and a web site creating a sense of community.

References

1. Shih, H.-P.: An Empirical Study on Predicting User Acceptance of E-shopping on the Web. *Information & Management* 41, 351–368 (2004)
2. Shih, H.-P.: Extended Technology Acceptance Model of Internet Utilization Behavior. *Information & Management* 41, 719–729 (2004)
3. Negash, S., Ryan, T., Igbaria, M.: Quality and Effectiveness in Web-based Customer Support Systems. *Information & Management* 40, 757–768 (2003)
4. Piccoli, G., Brohman, M.K., Watson, R.T., Parasuraman, A.: Net-based Customer Service Systems: Evolution and Revolution in Website Functionalities. *Decision Sciences Journal* 35, 423–455 (2004)
5. Zeithaml, V.A.: Service Quality, Profitability, and the Economic Worth of Customers: What we Know and What we Need to Learn. *Journal of Academy of Marketing Science* 28, 67–85 (2000)
6. Parasuraman, A., Zeithaml, V.A., Berry, L.L.: A Conceptual Model of Service Quality and its Implications for Future research. *Journal of Marketing* 49, 41–50 (1985)
7. Van Iwaarden, J., Van der Wiele, T., Ball, L., Millen, R.: Perceptions About the Quality of Web Sites: a Survey amongst Students at Northeastern University and Erasmus University. *Information & Management* 41, 947–959 (2004)

8. Caruana, A., Ewing, M.T., Ramaseshan, B.: Assessment of the Three-column Format SERVQUAL: An Experimental Approach. *Journal of Business Research* 49, 57–65 (2000)
9. Cook, C., Thompson, B.: Reliability and Validity of SERVQUAL Scores Used to Evaluate Perceptions of Library Service Quality. *The Journal of Academic Librarianship* 26, 248–258 (2000)
10. Lindgaard, G., Dudek, C.: What is this Evasive Beast we Call User Satisfaction? *Interacting with Computers* 15, 429–452 (2003)
11. Muylle, S., Moenaert, R., Despontin, M.: The Conceptualization and Empirical Validation of Web Site User Satisfaction. *Information & Management* 41, 543–560 (2004)
12. Yang, Z., Cai, S., Zhou, Z., Zhou, N.: Development and Validation of an Instrument to Measure User Perceived Service Quality of Information Presenting Web Portals. *Information & Management* 42, 575–589 (2005)
13. Garrity, E.J., Glassberg, B., Kim, Y.J., Sanders, G.L., Shin, S.K.: An Experimental Investigation of Web-based Information Systems Success in the Context of Electronic Commerce. *Decision Support Systems* 39, 485–503 (2005)
14. Li, Y.N., Tan, K.C., Xie, M.: Measuring Web-based Service Quality. *Total Quality Management* 13, 685–700 (2002)
15. Parasuraman, A., Zeithaml, V.A., Malhotra, A.: E-S-QUAL: a Multiple Item Scale for Assessing Electronic Service Quality. *Journal of Service Research* 7, 213–234 (2005)
16. Barnes, S., Vidgen, R.: An Integrative Approach to the Assessment of Ecommerce Quality. *The Journal of Electronic Commerce Research* 3, 114–127 (2002)
17. Barnes, S., Vidgen, R.: WebQual: An Exploration of Web-site Quality. In: *Proceedings of the Eighth European Conference on Information Systems* (2000), <http://is.lse.ac.uk/asp/aspectis/20000052.pdf>
18. Barnes, S., Vidgen, R.: Assessing the Quality of Auction Web Sites. In: *Proceedings of the Hawaii International Conference on Systems Sciences* (2001), <http://www.webqual.co.uk/papers/auction.pdf>
19. Barnes, S., Vidgen, R.: Interactive E-government: Evaluating the Web Site of the UK Inland Revenue. *Journal of Electronic Commerce in Organizations* 2, 42–63 (2003)
20. Barnes, S., Vidgen, R.: Data Triangulation in Action: Using Comment Analysis to Refine Web Quality Metrics. In: *Proceedings of the 13th European Conference on Information Systems* (2005), <http://www.webqual.co.uk/papers/comment.pdf>
21. Soydal, İ.: *Web Bilgi Sistemlerinde Hizmet Kalitesi (Service Quality of Web Information Systems)*. Unpublished doctoral dissertation, Hacettepe University, Ankara (2008)
22. Tonta, Y., Soydal, İ.: Service Quality of Web Information Systems. In: Katsirikou, A., Skiadas, C.H. (eds.) *Qualitative and Quantitative Methods in Libraries: Theory and Applications (Proceedings of the International Conference on QQML 2009)*, pp. 390–397. World Scientific, London (2010)
23. Barnes, S., Vidgen, R.: Measuring Web Site Quality Improvements: a Case Study of the Forum on Strategic Management Knowledge Exchange. *Industrial Management and Data Systems* 103, 297–309 (2003)
24. Heron, P., Calvert, P.: E-service Quality in Libraries: Exploring its Features and Simensions. *Library & Information Science Research* 27, 377–404 (2005)
25. Vavra, T.G.: *Improving your Measurement of Customer Satisfaction: A Guide to Creating, Conducting Analyzing, and Reporting Customer Satisfaction Measurement Programs*. ASQ Quality Press, Milwaukee (1997)
26. What is quadrant analysis?, <http://www.people-first.com/quad.html>
27. National Survey Research Center: Analyzing Customer Satisfaction Survey Results Using Gap Analysis, http://www.nsrc.com/services/ac_gap.html

Information Science and Cognitive Psychology: A Theoretical Approach

Armando Malheiro da Silva, Fernanda Ribeiro, and Fernanda Martins

Faculdade de Letras da Universidade do Porto / CETAC:MEDIA, Via Panorâmica,
s/n 4150-564 Porto, Portugal
armando.malheiro@gmail.com, fribeiro.hierro@netcabo.pt,
martins.f@clix.pt

Abstract. Information, as a human and social phenomenon, is the object of study of an emergent scientific field named Information Science (IS), which we put forward as unitary and transdisciplinary and open to a rich interdisciplinarity with other fields of knowledge. In face of the new reality, baptized the ‘Information Society’, and the emergence of a new paradigm, that we name “post-custodial, scientific and informational”, as opposed to the previous one, “historicist, custodial and technicist”, it is urgent to consolidate the theoretical and methodological foundations of IS in order to develop research, both pure and applied, and to contribute to a definition of its boundaries as a scientific area, in the scope of Social Sciences. Starting from an operative definition of Information, this paper aims to discuss the cognitive and emotional dimension of the info-communicational phenomenon and, for that, it is crucial to start a profound and hard dialogue with Cognitive Sciences. The label of ‘cognitivist’ given, in IS literature, to some authors like Bertram Brookes, because of the emphasis he put on the passage from a state of knowledge to a new state through an addition of knowledge coming from an increase of information, sounds quite equivocal, because knowledge and cognition are not synonymous and cognitive and emotional activity is not reducible to formalities. It is necessary to compare concepts and to understand the neuropsychological roots of the production, the organization and the info-communicational behaviour, so the contribution of Neurosciences and Cognitive Sciences, namely Cognitive Psychology, is indispensable.

Keywords: Information science, cognitive psychology, interdisciplinarity.

1 Introduction

In spite of the ancient roots of IS, it was only at the end of the 1950s that the term *Information Science* appeared in the literature, as a natural evolution from Documentation, boosted by the new Information and Communication technologies and by the development of scientific and technical information.

Therefore, IS growth occurred in a “straight line” from Documentation, with strong input from *special librarians* and *documentalists*, and some contributions from traditional librarians, but without true commitment from archivists and records managers

[1]. It has been a path, in great measure, followed by people in professions related to information and stimulated by professional associations, which gradually gained its own space in academic and scientific contexts.

According to Shera and Cleveland [2], the event that marked the shift from Documentation to IS was the International Conference on Scientific Information that took place in Washington in 1958, as a result of a partnership between the American Documentation Institute, the International Federation of Documentation, the National Academy of Sciences and the National Research Council. Thus, this event brought together the major players at world level.

Hans Wellisch states that the expression, Information Science, was used for the first time in 1959 and Anthony Debons [3] says that in 1962 the same expression appeared in the name of an international meeting, the Second International Congress on Information System Sciences, which took place in Hot Springs, Virginia (USA) [4]. In effect, by the 1970s, the expression enjoyed wide-spread acceptance in the USA and the area was more rapidly developed than in other countries. Thus proposals for a definition started to surface, as well as various articles aiming to establish the theoretical grounding of this new scientific field.

2 The Concept of Information and the Foundations of Information Science

Since the 1960s, multiple definitions have been advanced. However, one of the most accurate and complete came out during the Conferences of the Georgia Institute of Technology (Oct. 1961-Apr. 1962) and was further elaborated upon by Harold Borko in a seminal article titled Information Science - what is it? [5]. It remains, still today, one of the most consensual and fruitful:

Information Science is that discipline that investigates the properties and behavior of information, the forces governing the flow of information, and the means of processing information for optimum accessibility and usability. It is concerned with that body of knowledge relating to the origination, collection, organization, storage, retrieval, interpretation, transmission, transformation, and utilization of information. This includes the investigation of information representations in both natural and artificial systems, the use of codes for efficient message transmission, and the study of information processing devices and techniques such as computers and their programming systems. It is an interdisciplinary science derived from and related to such fields as mathematics, logic, linguistics, psychology, computer technology, operations research, the graphic arts, communications, library science, management, and other similar fields. It has both a pure science component, which inquires into the subject without regard to its application, and an applied science component, which develops services and products. (...) Librarianship and documentation are applied aspects of information science.

Looking at the history of IS leads to the realization that this field and the disciplines from which it derives began, above all, as a practical activity. This circumstance raises

some problems to the scientific assertion and identity of IS. A study by Emilio Delgado López-Cózar on the research conducted in Librarianship and Documentation focuses precisely on this problem and the author recognizes that:

The origin and evolution of disciplines of a professional basis, such as IS, cannot be explained through the same conceptual assumptions used in the sciences *per se*. The latter deepen their roots in the desire, innate to the human being, of knowing in and for itself, that is, to understand the world and to dominate it in benefit of the human species. They are born and grow in function of a dual human need: the cognoscitive (to know and to explain the why and the how of our natural, social and human reality) and the utilitarian (to apply the knowledge to the improvement of our life conditions). This conceptual scheme is valid to shape the history of all sciences, but particularly so of those which justify themselves only as a means to satisfy that dual need. They are sciences *per se*: the human and social sciences (philosophy, theology, history, sociology...) and the physical-natural sciences (physics, chemistry, biology...). But this is not the case when explaining the history of sciences that have appeared as an activity and as the support to a profession: from medicine or nursing to education, including a wide range of professions which arose from the economic, social and cultural development of humanity, such as social work or Librarianship and Documentation, which is what concerns us here. In these cases, their very character as a scientific profession is still under discussion. [6]

We absolutely agree with López-Cózar when he states that “in the development of Librarianship and Documentation [and we add Archivistics] theory followed practice, neither directing nor guiding it” [6]. In fact, the professional activity stimulated reflection on the *praxis* and, consequently, the need for some kind of training arose, appropriate to such a professional activity. Reflection on practice therefore led to disciplinary affirmation, to theoretical ‘exercises’ and to research work. These factors have been fundamental in building scientific knowledge around an object of study (information) and to boost the emergence of a science. But this scientific construction did not occur at the same time and in the same way in every country or context, which means that the degree of IS development is quite variable and reaching a consensus about this scientific field has proved quite difficult.

On the one hand, we must acknowledge that Archivistics has been neglected (or even self-neglected) and archivists have essentially been left out of the evolutionary process of IS, even though they try to show their scientific identity, albeit without a theoretical basis; on the other hand, there is no scientific consensus about the epistemological unity of the field, which could contribute to an understanding of IS as an interdisciplinary field; furthermore, some perspectives still persist, sustained by the traditional paradigm that considers only recorded information (=Documentation) as the object of study, which restrict the understanding of the informational phenomenon to an epiphenomenon, causing perverse effects from a scientific point of view.

In spite of this multiplicity of positions, we think that it is possible to contribute to clarifying this issue by defending IS as a unitary yet transdisciplinary field of knowledge, included in the overarching area of the human and social sciences, which gives theoretical support to some applied disciplines such as Librarianship, Archivistics,

Documentation and some aspects of Technological Information Systems. The way in which we see the cartography of the IS scientific field at the University of Porto is represented in a diagram, designed in 2002 [1], and later rebuilt [7], so as to better illustrate its transdisciplinary dimension and, at the same time, its interdisciplinary relationships (see Figure 1).

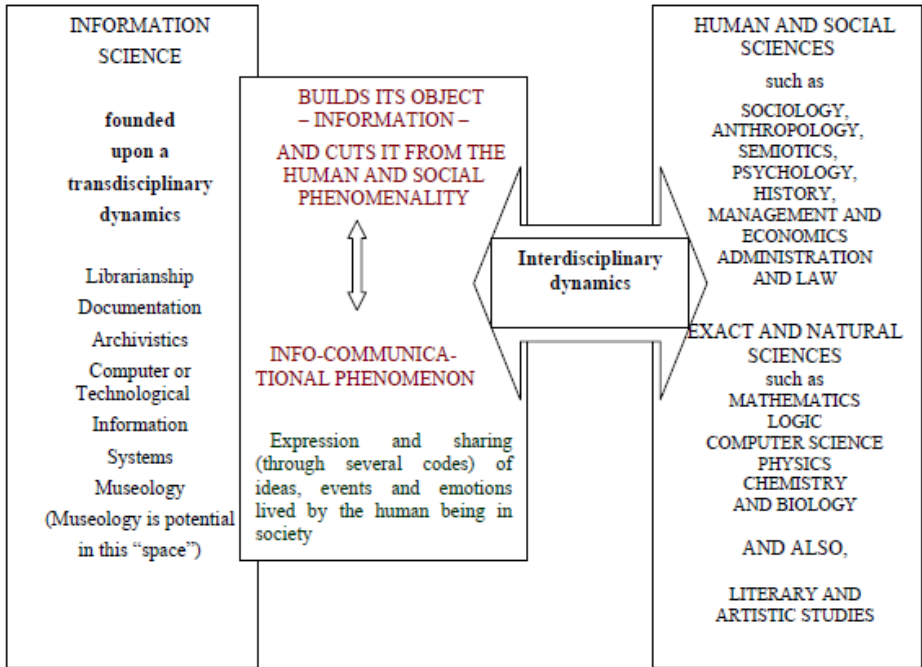


Fig. 1. Diagram of the trans- and interdisciplinary construction of Information Science

In this perspective, besides establishing the boundaries of IS, it is also crucial to define its object of study and to assume a research method adapted to the characteristics of Information as a social phenomenon, emphasizing its qualitative component, as is appropriate in the scope of the social sciences.

When it comes to IS's object of study and work – Information – it is essential to have a definition as a starting point, because it acts as an operative and foundational concept. The definition we propose is as follows:

Information is a structured set of mental and codified representations (significant symbols), created in a specific social context and capable of being recorded on any medium (paper, film, magnetic tape, compact disc, etc.) and, therefore, communicated in an asynchronous and multi-directed way [1].

Complementing the definition, the characterization of the informational phenomenon is broadened by the enunciation of its properties. In his book *A Ciência da Informação* [Information Science], Yves-François Le Coadic [8] attempted to formulate

the properties of information, but, in our opinion, in a way that is rather unclear. So, we attempt to complete the above definition by listing the *properties* of information, formalized as general axioms. Information is:

1. *structured by an action* (human and social) — the individual or societal act structurally establishes and models information
2. *integrated dynamically* — the informational act is involved with, and results from, conditions and circumstances both internal and external to that action
3. *has potentiality* — a statement (to a greater or lesser extent) of the act which founded and modelled the information is possible
4. *quantifiable* — linguistic, numeric or graphic codification is capable of quantification
5. *reproducible* — information can be reproduced without limit, enabling, therefore, its subsequent recording/memorization
6. *transmissible* — informational (re)production is potentially transmissible or communicable.

These six properties, and especially the last two, characterize information, not only as a *phenomenon* but also as a *process*. In this second dimension we include the idea of information behaviour, as well as all the activities related to the creation, organization, representation, storage, retrieval and use of information. So, information comprises the core (single and cross-disciplinary) of an academic field, which is itself dynamic and closely interrelated with other disciplines.

The assumption of social information as the object of knowledge has wide-ranging and unexpected implications. The main one is the emergence of a *scientific-informational* paradigm, shaped by the following factors:

- a) the value of information (and not the medium on which it is recorded) as a human and social phenomenon/process, with its own historicity (organic and contextual) and its cultural importance;
- b) the statement of the natural and continuous dynamism of information in opposition to documental immobility;
- c) the impossibility of keeping the traditional divisions of information according to the institutional or technological space where it is preserved (archival service, library or computer package) because such a criterion does not embrace the dynamic context of its production, of its recording and of its use/access (functionality);
- d) the need to know (to understand and to explain) social information through theoretical-scientific models, increasingly more effectively, instead of an empirical practice reduced to a set of technical procedures such as arrangement, description and retrieval;
- e) the replacement of the process-oriented perspective evident in the terms 'records management' or 'information management' by a new scientific view that tries to understand the information involved in the management process of any organization; this means that the informational practices/procedures are aligned with managers' conceptions and practices and with the organizational culture.

These characterizing elements, together with the definition of Information, can be considered the minimum and fundamental basis of a scientific approach to that which we consider to be the object of study and work of IS, understood as a theoretical and practical field in consolidation that supports multifaceted professional competencies, in accordance with the contexts and demands of professional activities.

In what concerns the methodological component of IS, we can sum up the ideas largely explored in the book mentioned previously [1]. According to the topological model proposed by Paul de Bruyne, J. Herman and M. de Schoutete for research in the social sciences [9], [10], the method of information science is achieving greater acceptance and tends to find consolidation through quadripolar research dynamics, which are operated and continuously repeated within the field of knowledge itself. This action combines quantitative approaches (there are aspects of the object which can be observed, experimented on and measured) and qualitative approaches, in which the subject’s interpretative/explanatory ability necessarily has modeling implications. The research dynamics mentioned thus imply permanent interaction on four poles, that is, the epistemological, theoretical, technical and morphological.

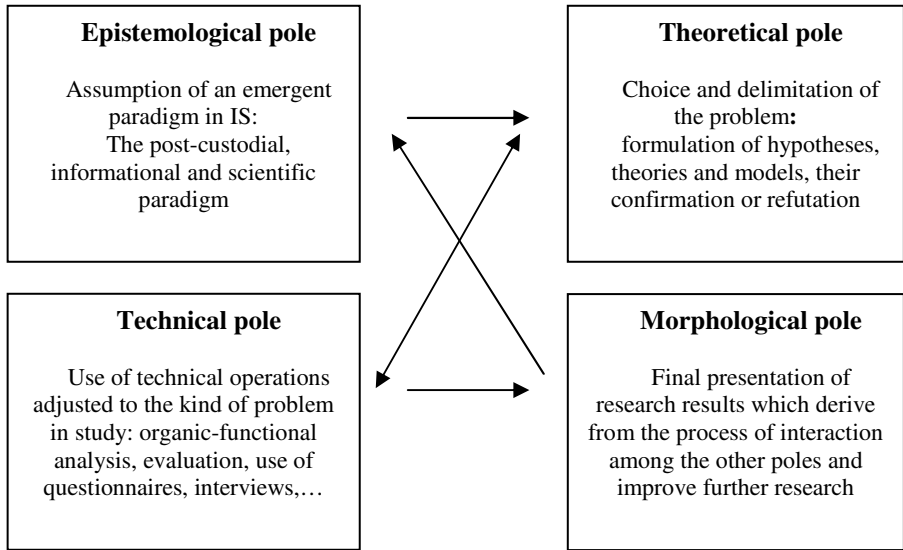


Fig. 2. Quadripolar method of research: interactions between the four poles

The *epistemological pole* — the scientific community of archivists, their schools, institutes, working places, with their own political, ideological and cultural references — operates the permanent construction of the scientific object and the definition of the boundaries of the research problems. The discursive parameters are constantly reformulated, as are the paradigms and scientific criteria (objectivity, reliability and evaluation) which guide the whole research process. Empirical procedures and archival knowledge gradually substantiate this pole, which is by no means static but, on the contrary, must be subject to periodic reflection on the occurrence, or otherwise, of epistemological continuity or gaps.

The *theoretical pole* operates the rationality of the subject (who knows and approaches) over the object, as well as the postulation of laws, the formulation of hypotheses, theories and operational concepts and the consequent validation or refutation of the “theoretical context” elaborated.

On the *technical pole*, contact with objectified reality is operated through instrumental application, thus verifying the validation capacity of the methodological mechanism. It is here that crucial operations are developed, such as the study of cases and variables and retrospective and prospective evaluation, always keeping in mind the confirmation or refutation of the postulated laws or principles, the theories elaborated and the operational concepts formulated.

On the *morphological pole*, the results of the research carried out are formalized through the representation of the object of study and the description of the whole research process which enabled the scientific construction around it. It deals with the organization and presentation of data, objectively checked on the theoretical and the epistemological poles, which shows the interactive character of the quadripolar method of research.

In this quadripolar dynamic, the theoretical pole assumes particular relevance, because it supports the technical and instrumental component and gives meaning to the explanation of the results in the morphological pole. There are, naturally, different theories and models applied to the interpretation of the informational phenomenon/process, but we prefer the Systemic Theory, whose origins derive from Ludwig von Bertalanffy’s studies, developed since the 1920s. This preference is based on the fact that Systemic Theory enables a holistic view and adjusts quite well to the complex and diffuse universe of Information [11].

This epistemological, theoretical and methodological foundation of IS, here briefly reviewed, is mirrored, obviously, in research projects, in educational and training models and in professional activities, developed in the most diverse organizational contexts and applies, obviously to interdisciplinary approaches with other scientific fields.

3 Some Contributions of Cognitive Psychology

During the last years Information Science, which is related with the production, conservation and communication of information, has changed in a way where the knowledge of the user’s cognitive functioning is a priority. The development of both Information and cognitive sciences makes it possible to better understand, for instance, memory and linguistic human functioning and use this knowledge to create powerful information systems related with new information and documentation practical issues.

In order to access complex and numerous information sources the users must have not only some special knowledge but also some abilities, namely for reading, searching, assessing and treating information. Underlying those actions there are the cognitive processes traditionally studied by Cognitive Psychology.

The first steps in Cognitive Psychology were made in order to understand the human psychological functioning, namely the mental processes, aspects completely ignored by Behaviorism, a radically different perspective. However, two different ways of studying these aspects appeared: connectionism, addressing particularly the

neural network and its states of equilibrium, and cognitivism (e.g. [12]). The concept of representation and the notion of process are fundamental in this last perspective. In fact, those mental representations are stable at a time, constituting our knowledge or they may be developed when necessary to a specific action. This is the way, for instance, that memory works updating and reconstructing knowledge as in the case of information seeking. Those representations have different forms and organization, namely declarative and procedural knowledge, schemata and so on.

The notion of process was inspired by computer functioning, meaning the treatment and transformation of information. Thus, human mental activity was supposed to work either with sequential, parallel or cascade treatment modalities having an automatic and/or controlled treatment of information. Studying the cognitive processing also meant to understand the processes involved in specific functions. This was the case of memory, reasoning and language which have been studied for a long time and are more and more well known. Memory is no longer seen as storage but as a dynamic system that detects and treats new information. Aspects as contexts and intentions are now being related with reasoning (e.g. [13]). Language, although much studied, needs probably to be seen differently by Cognitive Psychology, once it is mostly seen as an instrument for representing the world, and not as a system of communication as it is seen by Information Science.

Cognitive Psychology has mainly studied processes like perception, memory structure and functioning and knowledge organization. Nowadays it is also concerned, in a different perspective than cognitivism alone, with interaction situations where information becomes significant for individuals.

Besides those core processes, Cognitive Psychology also studies learning, problem solving, decision making and the processes involved in information seeking. Learning is an important aspect of information processing characterized by the ability to adaptively change behavior. It is also the capacity to understand contingencies between events and actions, a process that facilitates causal reasoning and induction leading to the development of categorization which organizes our knowledge [14]. Learning is behind judgments and decision making as they depend on prior experience and on related information that can be useful in the future [15]. Cognitive learning theories view learning as an individual construction and emphasize the active transformation of information needed to achieve changes in individual knowledge structures and to create personal meaning. Knowing about those aspects brings some light about the concept of information literacy which is an enabler of learning [16]. In fact, information literacy means to be able to process information and to reconstruct it meaningfully. It helps to identify whether the information found is relevant, an issue that is fundamental in the process of information searching [17], whether the source is authentic, and also to accept or reject the viewpoints encountered in various data sources. Although it mostly refers to individual processes, it is also related to social construction if attention is paid to the rapid expansion of collaborative learning, online learning communities and formal mechanisms for peer support.

Nowadays, the use of the Web as a channel of communication and as an important vehicle for information dissemination and retrieval has enlarged and improved information-seeking behavior [18]. However, seeking behavior studies, focused on traditional systems, are not able to provide sufficient information about users' interactions when searching the Web. They do not offer information about the different

information needs, cognitive and affective characteristics and experience of individual users. Research on Web information searching is useful for examining behavior and actions but is not adequate for explaining the factors and processes that have led to that behavior [19].

Another field of Cognitive Psychology is the study of emotions, important components of cognitive activity. Seen nowadays as generated by our cognitive evaluations of the environment, they are also important in the way they interact with cognitive processing (e.g. [20]). Studies that analyze, for instance, the relations between emotions and attention, memory and decision making either elicited by emotional stimuli or by individual emotional states. Obviously those emotional aspects are of great importance to Information Science, since information users are human beings processing information also emotionally. Information Science professionals are interested in knowing more about users' emotions, above all, about which emotions they should be concerned with. This is the case of knowing if the sources for information seeking and for decision making preferred by individuals affect their levels of satisfaction.

The developments of the humans' emotions knowledge underline most work on human-computer interaction that is currently being done. Although this area involves aspects like very technical issues, ergonomics, effectiveness and efficiency, it is also concerned with psychology and emotions particularly in many works in Information Science [21].

It is clear that users of technology products and interactive computing systems are no longer only interested in the product efficiency and effectiveness but that they are also looking for emotional satisfaction. Those emotional aspects are related with the trust in the service or product, the pleasure in the interactive experience, and the satisfaction with Web sites or products [22]. Positive emotions create positive affect leading to better decisions, and increase intrinsic motivation in people engaged in the activities.

The way emotions have been reached in most research studies in this area includes the use of behavioral indicators of frustration like false starts and input errors. Most often researchers use: (1) questions to users after the event; (2) verbal self-report; (3) users' responses to questions about satisfaction; (4) grids with semantically different words; or (5) standardized measuring instruments (e.g. [23]). However, more recently, some researchers, in order to assess emotions in a more authentic way, analyze the non-verbal aspects of emotional communication, particularly facial expression, based on the classical work of Paul Ekman (e.g. [24]). Emotional expression is videotaped for later observation and classification of the facial activity and facial expressions allowing easier recognition of the difficulties users may be feeling (e.g. [25], [26]). All those efforts are to achieve to users' satisfaction with the purpose of minimizing frustration, annoyance, anger and confusion.

Cognitive sciences are still concerned with the understanding of cognitive processes and use several methods to reach that goal, including traditional psychological experiments, observations of cognitive processing in practical action, or simulating cognition in robots or programs. Research is also increasingly connected with neuroscience and thus with new techniques of brain imaging allowing better understanding of brain functioning. Moreover the use of other techniques, like eye-tracking, will help enlarge our knowledge about the cognitive interaction with the world, and the practical applications of cognitive sciences will allow the creation of interfaces to information technology more and more adapted to the demands of human cognition [27].

4 Final Remarks

The conception of IS that is being developed at the University of Porto (Portugal) makes strategic and indispensable the construction of interdisciplinary approaches with several human and social sciences. What has been described previously clarifies themes and issues that can and must be deepened with the important contribution of Cognitive Psychology, namely through research on users' informational behaviour and over the complex processes of post-custodial mediation [28]. Thus, some promising fields of work have arisen, which have kept, in the meantime, a progressive interest and the attention of researchers involved in academic post-graduate programmes.

References

1. Silva, A.B., Ribeiro, F.: Das "Ciências" Documentais à Ciência da Informação: Ensaio Epistemológico Para um Novo Modelo Curricular. Porto, Edições Afrontamento (2002)
2. Shera, J.H., Cleveland, D.B.: History and Foundations of Information Science. *Annual Review of Information Science and Technology* 12, 249–275 (1977)
3. Debons, A.: Information Science. In: ALA World Encyclopedia of Library and Information Services, 2nd edn., Chicago, American Library Association (1986)
4. Lilley, D.B., Trice, R.W.: A History of Information Science: 1945-1985. Academic Press, San Diego (1989)
5. Borko, H.: Information Science - What is it? *American Documentation* 19, 3–5 (1968)
6. Delgado López-Cózar, E.: La Investigación en Biblioteconomía y Documentación. Ediciones Trea, Gijón (2002)
7. Silva, A.M.: A Informação: da Compreensão do Fenómeno e Construção do Objecto Científico. Porto, Edições Afrontamento, CETAC.COM (2006)
8. Le Coadic, Y.-F.: A Ciência da Informação. Trad. de Maria Yêda F. S. de Filgueiras Gomes. 2 edn. Briquet de Lemos – Livros, Brasília (2004)
9. De Bruyne, P., Herman, J., Schoutheete, M.: Dynamique de la Recherche en Sciences Sociales de Pôles de la Pratique Méthodologique. PUF, Paris (1974)
10. Lessard-Hèbert, M., Goyette, G., Boutin, G.: Investigação Qualitativa: Fundamentos e Práticas. Instituto Piaget, Lisboa (1994)
11. Mella, P.: Dai Sistemi al Pensiero Sistemico: Per Capire i Sistemi e Pensare Con i Sistemi. Franco Angeli, Milan (1997)
12. Ellis, H.C., Hunt, R.R.: Fundamentals of Cognitive Psychology. WCB Brown & Benchmark Publishers, Dubuque (1993)
13. Matlin, M.W.: Cognition. Fort Worth, Harcourt Brace Publishers (1994)
14. Newell, B.R., Bröder, A.: Cognitive Processes, Models and Metaphors in Decision Research. *Judgment and Decision Making* 3, 195–204 (2008)
15. Newell, B.R., Lagnado, D.A., Shanks, D.R.: Straight Choices: The Psychology of Decision Making. Hove, UK. Psychology Press (2007)
16. Markless, S.: A New Conception of Information Literacy for the Digital Learning Environment in Higher Education. *Nordic Journal of Information Literacy in Higher Education* 1, 25–40 (2009)
17. Saracevic, T.: Relevance: A Review of the Literature and a Framework for Thinking on the Notion in Information Science Part II: Nature and Manifestations of Relevance. *Journal of the American Society for Information Science and Technology* 58, 1915–1933 (2007)

18. Nahl, D.: Learning the Internet and the Structure of Information Behavior. *Journal of the American Society for Information Science* 49, 1017–1023 (1998)
19. Martzoukou, K.: A Review of Web Information Seeking Research: Considerations of Method and Foci of Interest. *Information Research* 10, paper 215 (2005)
20. Damásio, A.R.: *Descartes' error: Emotion, Reason and the Human Brain*. N Y, Hayrer Collins (1995)
21. Saracevic, T.: Interdisciplinary Nature of Information Science. *Ciência da Informação* 24, 1–9 (1995)
22. Shih, Y.-H., Liu, M.: The Importance of Emotional Usability. *Journal of Educational Technology Systems* 36, 203–218 (2008)
23. Brave, S., Nass, C.: Emotion in Human Computer Interaction. In: Jack, J.A., Sears, A. (eds.) *Human Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*. N.J., LEA (2003)
24. Ekman, P.: Facial Expression and Emotion. *American Psychologist* 48, 384–392 (1993)
25. Cristescu, I.: Emotions in Human-Computer Interaction: the Role of Nonverbal Behavior in Interactive Systems. *Informatica Economică* 2, 110–116 (2008)
26. De Lera, E., Garreta-Domingo, M.: Ten Emotion Heuristics: Guidelines for Assessing the User's Affective Dimension Easily and Cost-Effectively. In: *Proceedings of the 21st BCS HCI Group Conference HCI 2007*. Lancaster University, pp. 1-3 (2007)
27. Gärdenfors, P.: Cognitive Science: From Computers to Anthills as Models of Human Thought. *Human IT* 3, 9–36 (1999)
28. Silva, A.M.: *Mediações e Mediadores em Ciência da Informação*. *Prisma.Com* 9 (2009), <http://prisma.cetac.up.pt/>

Visualizing the Structure of Scientific Output of Iranian Scholars in Science Citation Index (SCI) during 2000-2006

Farideh Osareh and Maryam Keshvari

Shahid Chamran University, Faculty of Education and Psychology, Department of Library and Information Science, Ahwaz-Iran
fosareh@yahoo.com, ma.keshvari@gmail.com

Abstract. To visualize the structure of Iranian scientific output in Science Citation Index (SCI), accessible via Web of Science (WOS), during 2000-2006, we used scientometric techniques and HistCite software. The number of Iranian documents indexed in SCI during the study period was 24480. Generally HistCite analyzes citation data on two different levels: based on citations in WOS (Global Citation Scale), and citations in collection of retrieved documents (Local Citation Scale). The results of this study showed that, in the study period a total of 8 clusters have been formed on the two levels (GCS and LCS): Clusters 1 and 2 (with 3 sub-clusters) in GCS and clusters 3, 4, 5, 6, 7, and 8 in LCS. The subject area of whole clusters was chemistry, but different areas of this discipline. The prominent subject area in our study was organic chemistry. The most effective document in this study was an article by Zolfigol with 123 global citations and 71 local citations. The subject category of cluster 1 was analytical chemistry and membrane electrodes. Cluster 2 consists of 3 sub-clusters (sub-cluster 1 hydrocarbons, sub-cluster 2 in the field of oxidation and nitrogen, and sub-cluster 3 catalysts). Cluster 4 was crystal structure, cluster 5 electrochemical analysis, cluster 6 macro cycles, cluster 7 aliphatic and aromatic complexes and the 8th cluster was polymers..

Keywords: Scientometrics, historiographical map, research activities, scientific products, scientific output, Thomson ISI, citation indexes, Web of Science (WOS), Iran.

1 Introduction

The issue of scientific production was first introduced at the University of Tehran in 1978 but was not pursued. In 1993-1994 this topic was studied more seriously, and in 1997 the citation role in evaluating scientific collaborations was the main subject area of the "Conference of Methodology and Research Techniques" which was carried out by the research deputy of the University of Tehran [1]. After 1999, scientific production was considered by the Iranian Ministry of Science, Research and Technology (MSRT). According to the law passed by MSRT, each Iranian who publishes an article in a Thomson ISI Journals receives a considerable reward [2].

The government supports and encourages the attention and tendency of Iranian authors towards internationalized Iranian research output, and this has caused a gradual increase in Iranian scientific output [3].

A simple search in WOS on Iranian scientific output showed that an exponential increase has happened in recent years. For instance, the number of Iranian publications in SCI in 2000 was only 1371, while this rate increased to 5630 in 2005, 13440 in 2008 and 16492 in 2009. In this article, we decided to study Iranian scientific output using scientometric indicators.

2 Importance and Necessity of Research

One of the important factors for sustainable development in each country is the number of the scientific products indexed in international accredited databases. Based on the importance of scientific products on national and international scales, this paper is trying to visualize scientific output of Iranian scholars in WOS during 2000-2006. To this end, the subject areas of established clusters will be identified, and the effective authors and articles in both GCS and LCS will be recognized. We also try to explore the participation rate of Iranian scholars in creating these products. In other words, in this research we consider the scientific products of Iran in the Thomson ISI. We should draw your attention to the Iranian national language which is Farsi; therefore a lot of Iranian scientific output is in Farsi so does not have a chance to be indexed in Thomson ISI databases. Hence, in this research we have mapped Iranian non-Farsi publications in the Science Citation Index via WOS.

3 Aims and Purposes

The main purpose of this research is to visualize the structure of the scientific products of Iranian scholars indexed in Thomson ISI accessible via WOS during 2000-2006. According to this map, we will study the history of science and recognize the effective authors, based on citations on both LCS and GCS. To reach the above goals we should answer the following questions.

4 Research Questions

1. How is the historiographical map of Iranian scientific output in WOS during 2000-2006?
2. What are the most important scientific clusters formed in Iranian scientific output during 2000-2006? And what are their subject areas?
3. Who are the most productive and effective Iranian authors on both citation levels (GCS & LCS) during 2000-2006?
4. What are the most effective articles based on both GCS and LCS citation levels and their publication dates?

5 Background

Osareh and Wilson [3] analyzed international collaboration of Iranian scientific publications in SCI during 1995-1999. The results show that Iran's publication output in science and technology increased dramatically in the SCI during 1995-1999. One of the most important and significant factors that caused this rise seems to be the government's research policies in the last few years of their study. In 1996, the Iranian government announced the first national research call for papers and continued it for the following years. The researchers selected topics according to their areas and started working with large research grants. This can lead the researchers towards the research topics needed by the government. Another result of this study shows that Iran's main international collaborators are authors with institutional affiliations in the US or the UK. However, it is obvious that Iran is looking more and more for collaborative partners elsewhere. Collaboration with authors in Canadian and Australian institutions has increased either in absolute numbers, relative percentages or both.

Osareh and McCain [4] tried to draw the intellectual structure of Iranian chemistry research in Science Citation Index (SCI). The results of this research showed that since 1990, Iranian chemistry research, as represented in the SCI, has grown at a rate of roughly 26% and 7 major clusters, Oxidation of Organic Compounds, Physical Organic Chemistry, Ionosphere, Analytical Chemistry, Solvent-Free Synthesis, C.J. Pedersen and Crown Ethers, Synthesis of Carbonyl Compounds, were identified. The topic areas were primarily in organic chemistry, and secondarily in analytical chemistry; other major topic areas such as biochemistry, applied chemistry, and chemical engineering were not seen.

6 Research Methodology and Data Gathering

The research method for this study was the scientometric method. The population of this research comprises 24480 documents produced by Iranian authors, indexed in SCI during 2000-2006.

Data were gathered and analyzed in 3 steps using 3 tools. In the first step, data were extracted by using SCI and via WOS in plain text format. In the second step, data were recognized by ISI.exe software, and in the third step data were registered into an Excel spreadsheet and made ready for analysis.

To draw the historiographical map of Iranian scientific output in SCI, we used HistCite software. This software is a product of ISI. Its input contains plain text files extracted from WOS and its output contains a graphical image of scientific outputs [5].

6 Data Analysis

6.1 Describing Data

The analysis of the data revealed that there were totally 24480 documents produced and indexed in SCI by Iranian authors during 2000-2006. The publication year,

document type and the language of the documents were analyzed and are displayed in Table 1.

The publication rate of Iranian scientific products increased from 2000 to 2006: 5.85 percent of the total for the period (1417 documents) were published in 2000 and 24.74 percent (5995 documents) in 2006. In other words, Iran increased its annual productivity 4.23 fold from 2000-2006.

The results of the analysis of the type of documents showed that the documents are in 13 different formats. The most frequent format was the article with 21513 (87.88%) titles, followed by meeting abstracts with 2139 (8.74%) titles (Table 1).

As can be seen in Table 1, Iranian scientific products in SCI during 2000-2006, were published in five different languages. English with 24499 (99.88%) documents ranked top, followed by French and German with 18 (0.07 %) titles, and 10 (0.04 %) titles respectively.

Table 1. Iranian Scientific Products by Language, Format and Publication Year

Publication year			Type of documents			Language		
Year	N	%	Format	N	%	Language	N	%
2000	1,417	5.85	Article	21,513	87.88	English	24,449	99.88
2001	1,775	7.32	Meeting Abstract	2139	8.74	French	18	0.07
2002	2,411	9.95	Literature Criticism	336	1.37	German	10	0.04
2003	3,234	13.35	Review	209	0.85	Russian	2	0.01
2004	4,062	16.77	Editor Review	200	0.83	Italian	1	0
2005	5,335	22.02	Correction	61	0.25	Total	24,480	100.0
2006	5,995	24.74	News	17	0.07			
Total	24,229	100.0	Biography	2	0.01			
			Database Review	1	0			
			Reprint	1	0			
			Bibliography	1	0			
			Book Review	0	0			
			Software Review	0	0			
			Article	21,513	87.88			
			Total	24,480	100			

6.2 Historiographical Map

A historiographical map has been drawn based on two separate levels, using HistCite: 1. Global citations scale (GCS). 2. Local citations scale (LCS). For the GCS map, the data sample was based on 300 documents (nodes). For the LCS map, due to the high number of links, and to have a clear graph we, drew the map with only 200 nodes.

Analyzing the Clusters of Iranian Scientific Products in SCI during 2000-2006.

Because of the length of the map, we decided to divide it by clusters and identify the clusters one by one. The results of the research showed that on the GCS level there were only 2 clusters; cluster 2 had 3 sub-clusters due to the diversity of the subject areas. In the LCS map we observed 6 clusters. The subject fields of clusters in the GCS map were "membrane electrode" (cluster 1); the 3 sub-clusters of cluster 2 were in "operation on hydrocarbons", "oxidation and nitrogen" and "catalysts". In the LCS

map, clusters were in “operation in organic chemistry”, “crystal structure”, “electrochemical analysis”, “macro cycles”, “aliphatic & aromatic complexes” and “polymers”. Firouzabadi and Heravy each participated in 3 clusters of which two had the same subject areas: “operation in hydrocarbons” and “operations in organic chemistry”. The subject category of the 3rd cluster for Firouzabadi was “aliphatic & aromatic complexes”, while Heravy participated in “catalysts” as the 3rd cluster. It should be noted that the first authors in all clusters were Iranian.

Scientific Clusters in SCI on GCS Map. Considering the GCS map, 2 clusters have been observed in Iranian scientific output in SCI during 2000-2006. Clusters 1 and 2 will be defined shortly. It should be noted that due to the subject diversity of cluster 2, it has formed 3 sub-clusters.

Cluster 1. This cluster has been established by the collaboration of 7 Iranian authors (Shamsipour, Ganjali, Mousavi, Shahrokhian, Javanbakht, Mashhadizadeh and Bagheri) during 1999-2005, in the area of “membrane electrode” in analytical chemistry. Figure 1 shows cluster 1, with the top 5 articles based on the number of citations and links which are shown in bold numbers in this Figure.

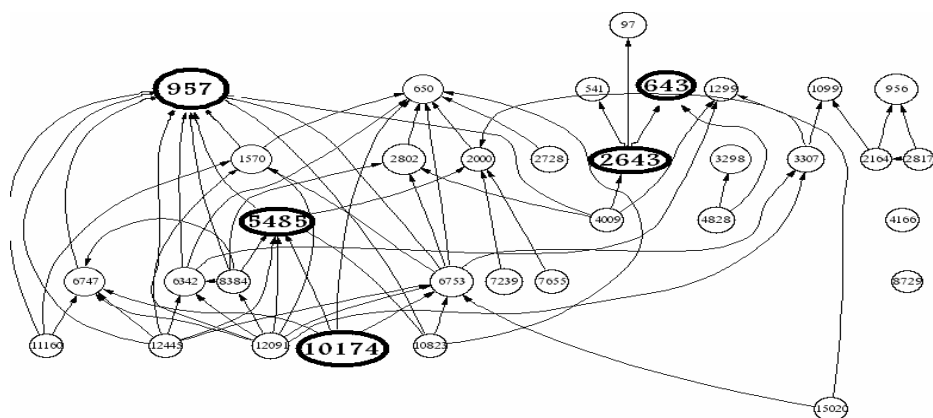


Fig. 1. Cluster 1 in membrane electrode area on GCS map

In this cluster, the most effective document (considering the number of citations received and links to it) is by Shamsipour (957) with 102 GCS. This document with 66 LCS is also the most effective document in LCS map.

Cluster 2. As was mentioned and can be seen in Figure 2, cluster 2 is a large and separable cluster. Therefore, it was divided in 3 sub-clusters due to the diversity of subject areas. Each sub-cluster will be analyzed separately.

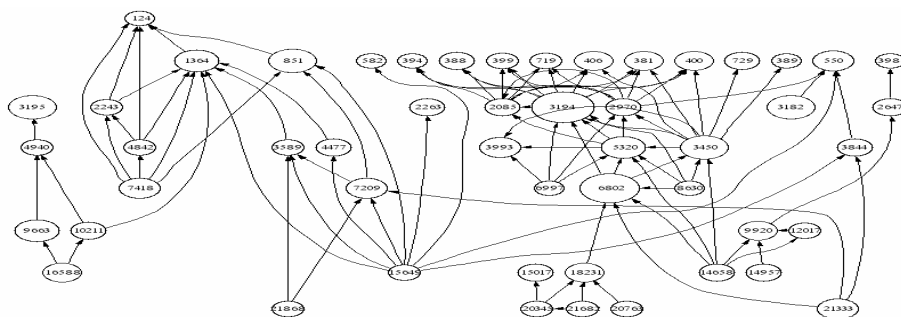


Fig. 2. Cluster 2 on GCS map

Sub-cluster 1 from Cluster 2. Figure 3 shows the first sub-cluster of cluster 2. This sub-cluster is established by collaboration of 7 Iranian authors (Kaboudin, Karimi, Azizi, Firouzabadi, Habibi, Saidi, and Heravy) during 1999-2006. The subject area of this cluster is "operation on hydrocarbons". The most effective document (3195) is by Kaboudin with 76 GCS. In Figure 3 the top 5 most effective documents in this sub-cluster are in bold.

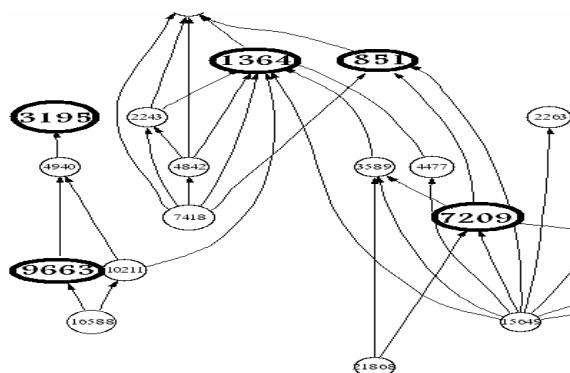


Fig. 3. Sub-cluster 1 from cluster 2 in operation on hydrocarbons on GCS

Sub-cluster 2 from Cluster 2. The subject area of this sub-cluster is "oxidation, nitrogen and catalysts" in the area of organic chemistry. The most important document in this sub-cluster is indicated by the number 3194 in Figure 4 and titled "Silica sulfuric acid/ NaNO_2 as a novel heterogeneous system for production of thionitrites and disulfides under mild conditions" by Zolfigol (2001). It received 123 GCS and 71 LCS. This document received the highest number of citations in both parts: GCS and LCS.

As can be seen in Figure 4 sub-cluster 2 is a big sub-cluster and has been established by collaboration of 10 authors (Zolfigol, Salehi, Firouzabadi, Shirini, Heravy, Khosropur, Mohamadpour, Balterak, Khodayi, Iranpour, and Sadeghi) by 28 documents during 2000-2006. Zolfigol with 14 documents produced the most articles in this sub-cluster. All of the first authors in these documents are Iranians. The top 5 most important documents from this sub-cluster are bold in Figure 4.

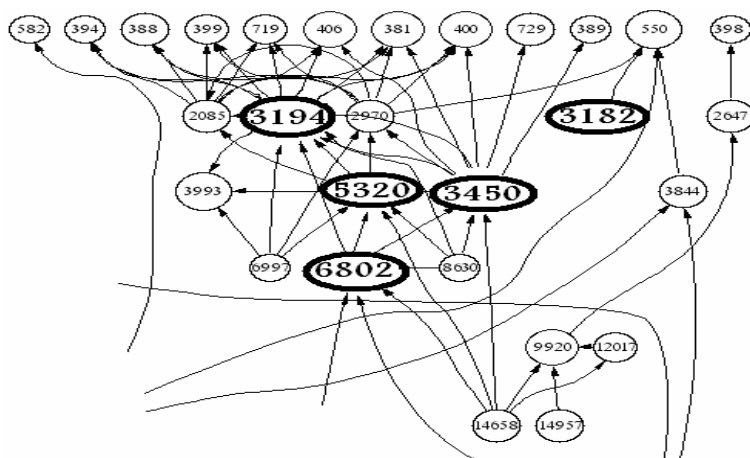


Fig. 4. Sub-cluster 2 of cluster 2 on GCS map

Sub-cluster 3 from Cluster 2. This is a small sub-cluster with only 5 documents and by collaboration of 2 Iranian authors (Heravy and Bamoharram). It was established during 2005-2006 in the "Catalysts" subject area. The most effective document in this cluster is document number 18231 from Heravy with 47 GCS.

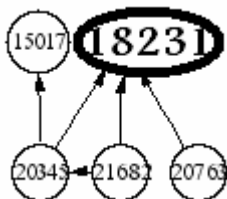


Fig. 5. Sub-cluster 3 from cluster 2 on GCS map

Clusters in SCI in Local Citation Scale. According to local scales, there are 6 clusters in scientific products of Iranian authors in SCI during the research period. Here we analyze them.

Cluster Number 3. This is another big cluster with 29 documents and collaboration of 8 authors (Zolfigol, Shirini, Salehi, Firouzabadi, Sadeqi, Keypour, Mirjalili, and Heravy). This cluster was established during 2000-2005 based on LCS. Zolfigol produced 21 documents out of 29 in this cluster. As was mentioned his document number 3194 received the most number of citations based on LCS and GCS. The subject category of this cluster is "oxidation, nitrogen and catalysts" in Organic Chemistry. First authors in all documents of this cluster are Iranian.

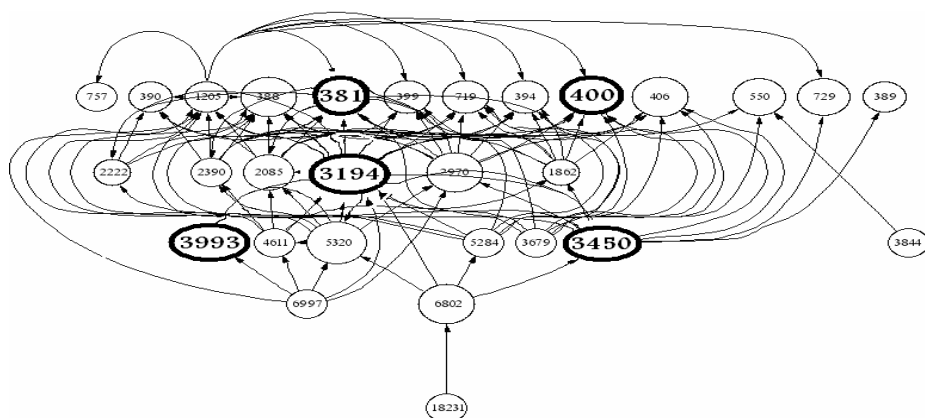


Fig. 6. Cluster 3 in oxidation, nitrogen and catalysts on GCS map

Cluster 4. This cluster has 9 documents produced by 2 authors (Moghimi and Ranjbar). It was established during 2001-2005 and its subject area is "crystal structure". The most effective document in this cluster is document 4170, by Moghimi, with 28 LCS.

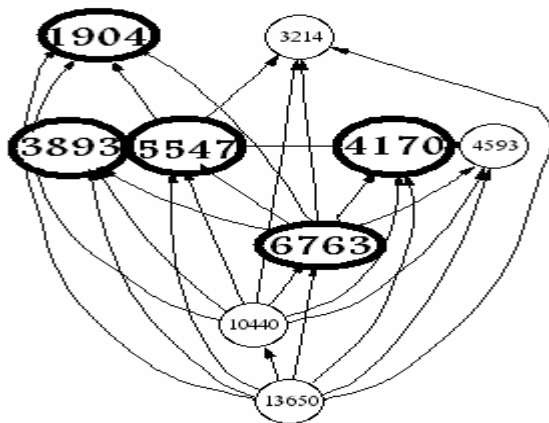


Fig. 7. Cluster 4 in Crystal structure on LCS map

Cluster 5. The subject area of this cluster is "electrochemical analysis". It has 29 documents with collaboration of 7 Iranian authors (Shamsipour, Shahrokhian, Ganjali, Javanbakht, Musavi, Rahmani and Bagheri). This cluster was established during 1999-2004 based on LCS.



Fig. 8. Cluster 5 in Electrochemical analysis based on LCS map

Cluster 6. This cluster has 7 documents by one author (Salavati Niasri). It was in "macro cycles" and has been established during 2003-2005. The most effective documents in this cluster are numbers 10714 and 11621 which each received 21 LCS.

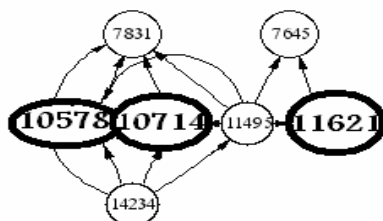


Fig. 9. Cluster 6 in Macro cycles on LCS map

Cluster 7. This cluster has 9 documents by 4 authors (Karimi, Firouzabadi, Azizi and Saidi) and was established during 1999-2004 in the area of aliphatic and aromatic complexes. The most effective document in this cluster is document number 1364 with 22 citations in local scale. The information of this cluster is available in Figure 10. All of the first authors in this cluster are from Iran.

Cluster 8. Cluster number 8 was established by 17 documents by 2 Iranian authors (Hajipour and Malekpour) during 2000-2004, in the subject area "Polymers". The most effective documents in this cluster are documents number 868 with 54, and documents number 558 and 852 each with 52 LCS.

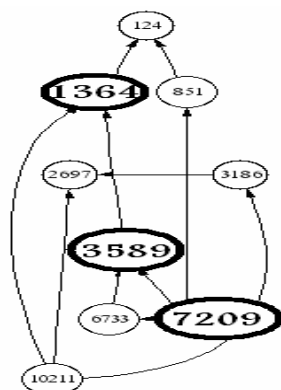


Fig. 10. Cluster 7 in aliphatic and aromatic on LCS map

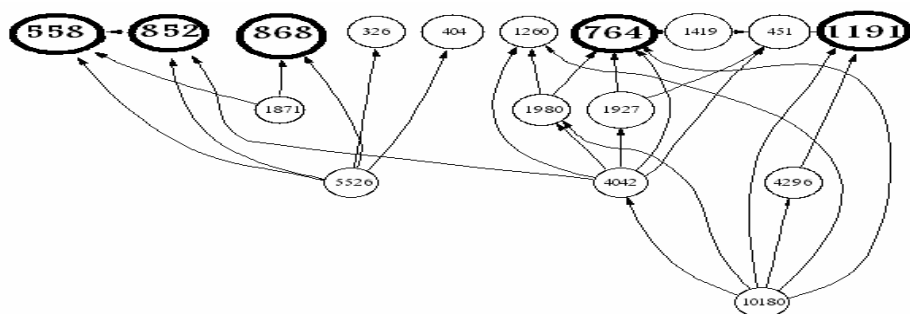


Fig. 11. Cluster 8 in polymers on LCS map

7 Conclusions and Suggestions

This article found that, the total rate of scientific production in the period of this research was increasing and the scientific products of Iran from 2000 to 2006 showed a growth rate of 4.23 percent. The most similar research to this study was done by Osareh and McCain [4] in which all clusters were about chemistry. These results were proved in our research and shown that generally, Iranian chemists have a tendency to produce scientific documents first in the field of organic chemistry and second in analytical chemistry. The most participant authors in the clusters of this study were Firouzabadi and Heravy who each participated in 3 clusters. Of those one was in the 3rd sub-cluster of cluster 2 (in GCS) and 2 other clusters in LCS. Among all 8 clusters, 1 cluster was formed by only 1 author. Two clusters were formed by publications of 2 authors. Other clusters had more than 2 authors. The results of this study also showed the influential Iranian authors and articles during 2000-2006 in WOS. The subject areas of clusters were recognized as follows:

“membrane electrode”, “operation on hydrocarbons”, “operation oxidation and nitrogen”, “organic chemistry” and “catalysts”. In LCS map, clusters were in “operation

in organic chemistry”, “crystal structure”, “electrochemical analysis”, “macro cycles”, “aliphatic & aromatic complexes” and “polymers”.

Using information visualization in different scientific disciplines could be useful for specialists as well as policy makers. The specialists at a glance can see which subject areas in their discipline have been under research by their colleagues, and which areas have been less attendant during a specific time. The results of such studies would let the policy makers allocate the budgets to subject fields with more confidence. The results of such studies also would be helpful for scientists and young specialists who can save their time by reading the works of key authors and influential scientific output in their disciplines.

References

1. Osareh, F.: Bibliometrics. *Journal of School of Education & Psychology of Shahid Chamran University* 8, 90–97 (1997)
2. Moosavi Movahedi, A.A., Kiani Bakhtiari, A., Khan Chamani, J.: Methods of Production and Dissemination of Scientific Findings. *Rahyaft* 31, 5–19 (2003)
3. Osareh, F., Wilson, C.S.: Collaboration in Iranian Scientific Publications. *Libri* 52, 88–98 (2002)
4. Osareh, F., McCain, K.W.: The Structure of Iranian Chemistry Research, 1990–2006: An Author Cocitation Analysis. *Journal of the American Society for Information Science and Technology* 59, 2146–2155 (2008)
5. Garfield, E., Paris, S., Stock, W.G.: HistCitedTM: A Software Tool for Informatic Analysis of Citation Linkage. *Information Wissenschaft und Praxis* 57, 391–400 (2006)

A Study on Scientific Product of the University of Tehran in Web of Science Database during 1989 – 2009

Farideh Osareh and Amin Zare

Shahid Chamran University, Faculty of Education and Psychology, Department of Library and Information Science, Ahwaz, Iran
{fosareh, azare}@scu.ac.ir

Abstract. The University of Tehran (UT) is the oldest and one of the largest academic centers in Iran. It plays an important role in education and research in the country. Due to the importance of Tehran University we decided to study its academic scientific output in the Web of Science during 1989-2009 using a scientometric approach. The purpose of this study was to identify the rank of UT among Iranian universities. The key authors and influential journals, types of documents, the rate of yearly output and the annual growth rate were also identified. We also specified the countries with whose scholars UT academic members collaborated during the studied period. In addition, we drew and analyzed two historiographical maps of UT, based on Local Citation Score (LCS) and Global Citation Score (GCS).

Keywords: University of Tehran (UT), scientific output, Web of Science, historiographical map, mapping of science.

1 Introduction

Nowadays scientific output, among others, is taken as one of the important criteria in the evaluating and ranking of different countries. Toffler [1] in his famous work titled *Powershift: knowledge, wealth, and violence at the edge of the 21st century* assumed that knowledge is power. In a similar vein, more than a thousand years ago, the Persian poet Ferdowsi maintained that “knowledgeable people are powerful”. As is the case at the present time, countries are evaluated not only by their national products, military power, geographical area, etc, but by such factors as the production and consumption of scientific information.

Along with the aforementioned issues, the production of scientific information has an increasing impetus in the era in which we live. Given this increasing speed as well as the huge volume of the produced knowledge and the plurality of scientific branches, one cannot examine all the scientific literature, even in a specific field, since managing huge amounts of information is a demanding task. Put another way, we can say that the dream of a comprehensive library or information center has not yet come true.

All of these discussions support the exploitation of citation index and databases as necessary measures for scholars and researchers in the field. In fact, these databases,

using quantitative methods and bibliometric and scientometric approaches, can be exploited as proper tools for judging scientific products at local, national and international levels.

In ranking universities, the number of documents recorded in databases is considered. However, focus on scientific products as the only measure is not reasonable, although we cannot deny their importance as good criteria for evaluating the scientific work of a person or an organization.

2 Problem Statement

The University of Tehran (UT), as the oldest and one of the largest universities in Iran, has an important role in higher education and research in the country. With a glance at the ranking of universities, we see that UT has the highest rank among Iranian universities [2]. Therefore studying the UT's scientific productivity as a method for evaluating its performance is reasonable. This research was intended to investigate the key authors from UT who have published in influential documents and journals included in the Web of Science 1989-2009. The main subject categories of the clusters in both UT historiographical maps during the studied period were also analyzed.

3 Goals and Questions

The present study evaluated the scientific productivity of UT, as indexed in WOS during 1989-2009. Besides, Iranian key authors, influential documents, yearly output and growth rate of UT scientific output were investigated in the study. To do so, we drew two historiographical maps of science based on GCS¹ and LCS² for UT. To reach the above goals, the following questions were raised:

1. What is the rank of UT according to scientific output compared to other Iranian universities during 1989-2009, in the WOS?
2. Who are the most productive authors in UT based on scientific output?
3. What are the types of documents written by UT academic members?
4. What are the languages of UT publications in WOS during 1989-2009?
5. What are the most important journals in which the studied documents were published?
6. What is the annual average rate of document production of UT academic members?
7. What is the annual growth rate of UT documents in WOS during 1989-2009?
8. What are the countries whose academics have the most frequent co-authorship with UT academic members?
9. How many clusters are included in the historiographical maps of UT?
10. What are the subject categories of the historiographical maps of UT?

¹ GCS - Global Citation Score shows the total number of citations to a paper in the Web of Science.

² LCS - Local Citation Score shows the count of citations to a paper within the collection.

4 Methodology

Applying the scientometric method, this study gathered data for UT scientific output by searching WOS on 24 January 2009. The result was 6099 records which had been published by at least one author affiliated to UT. For data analysis, HistCite™ and MS Excel were utilized. Data were extracted using analysis tools of WOS in some 500 sets. All records in 500 sets were entered in HistCite™. MS Excel was applied for drawing tables and figures.

5 Literature Review

Metrics methods have already been used by many researchers, while bibliometrics preceded other methods like scientometrics, webometrics and informetrics and go back some decades.

Osareh and Wilson [3] in their research of the collaboration on Iranian scientific publication, investigated three 5-year periods: 1985-1989, 1990-1994 and 1995-1999. They found that Iranian scientific publication in the second period was twice as much as in the first period, while in the third period it was about three times (2.8) more than in the second period.

Jacobs and Pichappan [4] investigated the scientific products of some universities in South Africa in ISI (Thomson Reuters) during 1994-2003. The results of their research showed that clinical science is the most productive field in scientific information produced by South African universities.

Lucio-Arias and Leydesdorff [5] pointed out the advantages of HistCite™ in drawing historiographical maps.

Osareh and McCain [6], in their article titled "The Structure of Iranian Chemistry Research, 1990-2006: An Author Co-citation Analysis", studied Iranian chemistry research. Their results revealed that the yearly growth rate of chemistry publication among Iranian authors was 26%. By using the method of co-citation analysis, they also introduced important factors in scientific products of Iranian chemistry.

In Iran, HistCite™ was introduced and used for the first time by Asnafi, Hamidi and Osareh [7]. They investigated scientific publications in the fields of Bibliometrics, Scientometrics, Informetrics and Webometrics in WOS during 1990-2005. They found that among 53 countries which collaborated in writing documents in the mentioned areas, the US ranked first followed by the UK, Germany and the Netherlands.

6 Data Analysis

The 6099 records gathered from WOS, were all published by authors affiliated to UT including faculty members and postgraduate students. All records were analyzed using the HistCite™ software. We can explain the results of data analysis as below:

According to WOS records, UT ranked first among Iranian universities. Table 1 shows the status of the top 10 universities of Iran according to WOS records. As can be seen, UT, with a total of 6099 records, has produced the most scientific documents followed by the Tehran University of Medical Science, Sharif University, Shiraz University and Tarbiat Modarres University.

Table 1. Top 10 Iranian Universities according to WOS records

#	Institute	Records
1	University of Tehran	6099
2	Tehran Univ of Med Sci	3800
3	Sharif Univ	3591
4	Shiraz Univ	3332
5	Tarbiat Modarres Univ	3115
6	Islamic Azad Univ	2398
7	Amir Kabir Univ	2334
8	Shahid Beheshti Univ of Med Sci	1900
9	Isfahan Univ of Technology	1829
10	Shiraz Univ of Med Sci	1501

Data analysis showed that among authors affiliated to UT, Ganjali with 300 records was the most productive author and ranked first. Mousavi Movahedi, Saboury, Norouzi and Zarrindast ranked 2-5 respectively.

Table 2 shows UT academic members ranked by the number of their publications. As can be seen in this table, authors with at least 40 publications are listed. In table 2, we can see that of the 6099 documents written by UT authors, 1982 documents (32.4%) were published by only 24 authors (0.25%). The 6099 documents had a total of 9400 authors, an average 1.5 authors for each document.

Considering type of the documents, 82.7% (5047 of 6099) were articles, followed by meeting abstracts and proceedings papers and nearly all documents [6079 of 6099= 99.6%] were written in English.

Table 2. Top 24 authors affiliated to UT with at least 40 documents in WOS during 1989-2009

Rank	Author	Records	Rank	Author	Records
1	Ganjali MR	300	13	Mahmudi R	50
2	Mousavi-Movahedi AA	251	14	Soltanian-Zadeh H	50
3	Saboury AA	231	15	Alimohammadi M	47
4	Norouzi P	174	16	Mohajerzadeh S	47
5	Zarrindast MR	98	17	Sarbolouki MN	47
6	Shamsipur M	84	18	Hakimelahi GH	46
7	Lucas C	69	19	Nemat-Gorgani M	45
8	Yazdanparast R	68	20	Ghandi M	44
9	Salavati-Niasari M	67	21	Darafsheh MR	43
10	Faiz J	60	22	Siavoshi F	41
11	Adib M	58	23	Yassemi S	41
12	Larijani B	51	24	Dehpour AR	40

These UT publications were published in 1926 journals. The number of journals that had published at least 5 UT documents was 333. The total number of documents published in these 333 journals was 3403. Thus of 6099 documents, 55.7% were published in 17.2% of all journals. According to Table 3, FEBS Journal published the most UT documents followed by Biophysical Journal and the Journal of Applied Polymer Science. Table 3 shows 10 top journals in which UT publications were published.

Table 3. Important journals in which UT documents have been published (first 10)

#	Journal	Articles
1	FEBS Journal	88
2	Biophysical Journal	54
3	Journal of Applied Polymer Science	49
4	Iranian Journal of Chemistry Chemical Engineering - International English Edition	45
5	Journal of Materials Processing Technology	42
6	International Journal of Psychology	36
7	Materials Science and Engineering A-Structural Materials Properties Microstructure and Processing	36
8	International Journal of Environmental Research	35
9	Applied Mathematics and Computation	34
10	Iranian Journal of Public Health	34

Another important issue was the production of scientific documents per year. 1991 was the only year in which the number of scientific documents decreased. In all other years they increased compared to the previous year. For more information about yearly output, see Table 4.

We also calculated yearly growth rate for publications of UT in WOS from 1989-2008. The number of UT publications grew at approximately 37.8% per year.

Table 4. UT publications in WOS by year

Year	Articles	Year	Articles
1989	3	1999	119
1990	11	2000	166
1991	6	2001	224
1992	32	2002	228
1993	41	2003	358
1994	46	2004	479
1995	48	2005	592
1996	80	2006	857
1997	89	2007	1278
1998	94	2008	1333

The authors of UT publications collaborated with colleagues from 81 countries. Among them, were 411 from the USA, followed by Canada, UK, Germany and France. Table 5 shows the collaboration between authors of UT and other countries.

Table 5. Collaboration between authors of UT and other countries (first 10)

#	Country	# of Co-Works
1	USA	411
2	Canada	250
3	UK	184
4	Germany	151
5	France	94
6	Japan	85
7	Australia	68
8	Italy	53
9	Taiwan	51
10	Sweden	43

There are some tools which can be used in drawing the structure of science in each field. For example, we can use SPSS, PathFinder and more recently HistCite™. Among them, HistCite™ has more capability in drawing the map of science and the structure of a field, like its ability to provide detailed information about authors, journals, cited references, keywords, yearly output and other data. Added to these applications, HistCite™ can draw historiographs based on Local Citation Score (LCS) and Global Citation Score (GCS) to show the important works and history of science in a field or in an organization, so we decided to use HistCite™ for this study.

Our study showed that the 6099 UT documents received 6621 local citations and 19,562 global citations. We extracted 60 top documents to draw two separate graphs based on LCS and GCS. These 60 documents received a minimum of 19 and a maximum of 60 local citations.

Records 243³ and 247⁴ were the first documents which received local citations in 1996. Record 654⁵ received the most local citations in comparison with the other 60 top documents, followed by 1076⁶, 451⁷, 330⁸ and 1336.⁹ These are important documents and the focal point of the main cluster.

As we can see in Figure 1, each circle is an indicator for a document: the larger the circle the more citations.

³ Saboury AA, 1996, J CHEM THERMODYN, V28, P1077.

⁴ Saboury AA, 1996, BULL CHEM SOC JPN, V69, P3031.

⁵ Shamsipur M, 2000, ANAL CHEM, V72, P2391.

⁶ Ganjali MR, 2004, SENSOR ACTUATOR B-CHEM, V98, P92.

⁷ Ganjali MR, 1998, ANAL CHEM, V70, P5259.

⁸ Fakhari AR, 1997, ANAL CHEM, V69, P3693.

⁹ Ganjali MR, 2003, TALANTA, V59, P613.

All of the mentioned articles which are the core of the main cluster were published by faculty members of the Department of Chemistry. We can see a type of collaboration among colleagues in this area. This graph also showed some self-citations and mutual citations. Of course, because of the similarity of their area, this type of citation behavior is natural. We saw two clusters besides the main cluster in the graph of LCS.

The cluster located at the left of the main cluster belonged to authors affiliated with the Institute of Biochemistry and Biophysics of UT. We can also observe a type of collaboration among some colleagues with the same research interest and methods, who are working in nearby offices.

There is another cluster located to the right of the main cluster. This cluster belonged to two colleagues in the Chemistry Department, but was separated from the main cluster and made an isolated cluster.

This graph shows that according to LCS, authors in the fields of Chemistry and Biochemistry & Biophysics are the cornerstone of science structure in UT. According to this graph Ganjali with 29 articles (of the 60 top articles) played a very important role.

We drew another graph with the 60 top documents based on GCS. Totally these 60 top documents received a minimum of 33 and a maximum of 166 global citations. We can see in the graph that among the 60 top documents, record 103¹⁰ was the first document which received global citations in 1994, followed by records 166¹¹ and 180¹² which received global citations in 1995. See Figure 2.

According to this graph, record 330¹³ with 166 global citation occupied the first rank, followed by records 654¹⁴, 680¹⁵, 481¹⁶ and 451¹⁷ with, respectively, 106, 92, 86 and 74 global citations. The GCS graph included three clusters. Chemistry, biochemistry and biophysics were the main part of the GCS graph.

This showed that from the viewpoint of global citation, authors from the Chemistry Department and the Institute of Biochemistry and Biophysics of UT played the most important role in the producing of science at UT. Thus, with respect to participation of scientific fields, there is no significant difference between the LCS and GCS graphs.

We also saw the same citation behavior in the GCS graph as in the LCS graph as a result of collaboration among colleagues, self and mutual citation and similarity of scientific areas. According to GCS, Ganjali was the most productive author. He, with 21 articles in the GCS graph (of the 60 top documents) played an important role. Here we can see the similarity between the two graphs.

¹⁰ Khosravi AR, 1994, MYCOSES, V37, P43.

¹¹ Vetter W, 1995, CHEMOSPHERE, V30, P1685.

¹² Testillano PS, 1995, EXP CELL RES, V221, P41.

¹³ Fakhari AR, 1997, ANAL CHEM, V69, P3693.

¹⁴ Shamsipur M, 2000, ANAL CHEM, V72, P2391.

¹⁵ Zimmer S, 2000, J BIOL CHEM, V275, P25672.

¹⁶ Javanbakht M, 1999, ELECTROANAL, V11, P81.

¹⁷ Ganjali MR, 1998, ANAL CHEM, V70, P5259.

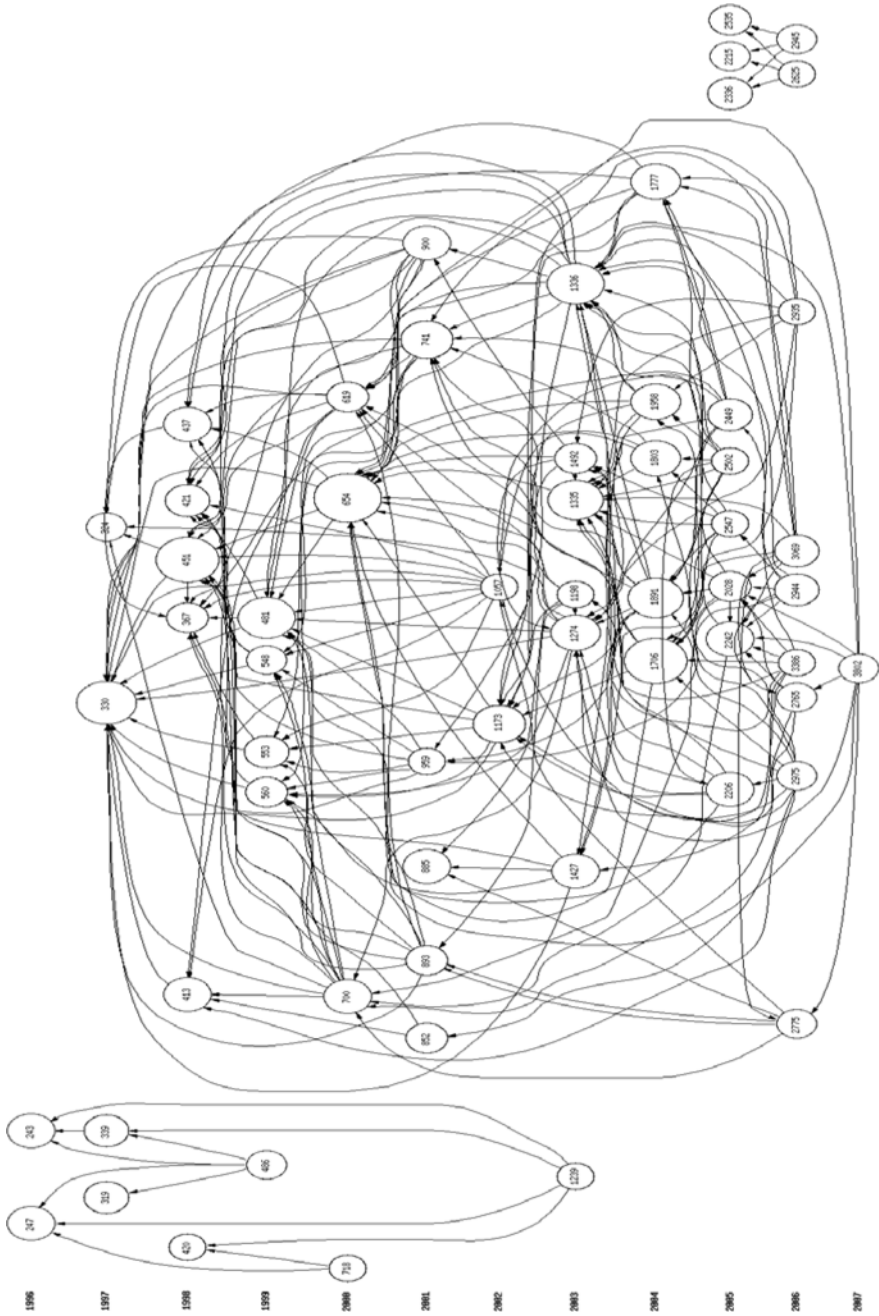


Fig. 1. Historiograph of UT based on LCS with 60 top documents

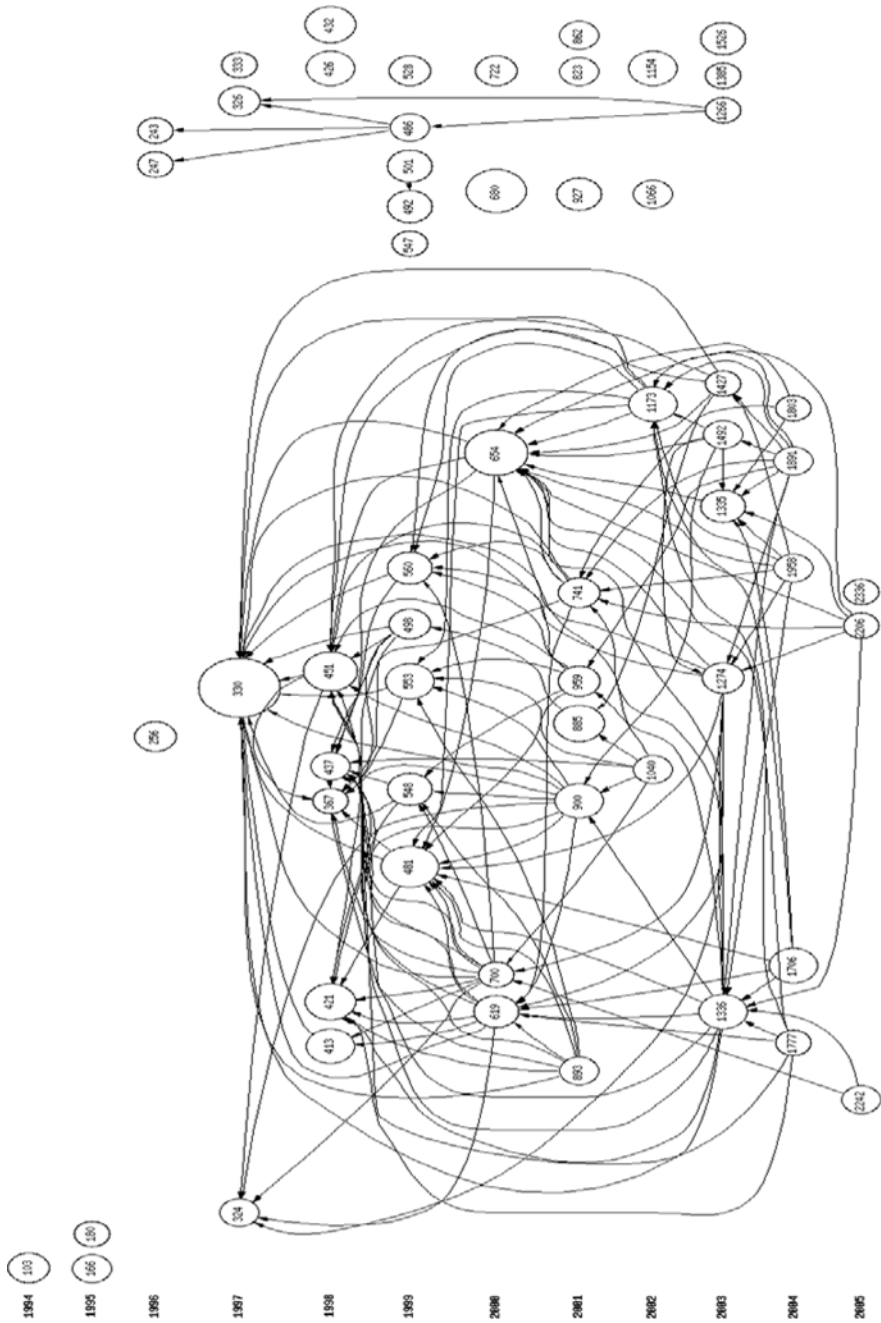


Fig. 2. Historiograph of UT based on GCS with 60 top documents

7 Conclusion

The results of this study showed that UT authors published 6099 scientific documents during 1989-2008. According to the number of publications extracted from WOS, UT published more documents than other Iranian universities and ranked first among them followed by Tehran University of Medical Science, Sharif University, Shiraz University and Tarbiat Modarres University.

Ganjali, from the Department of Chemistry, was the most productive author (with 300 articles) followed by Mousavi Movahedi, Saboury, Norouzi and Zarrindast.

Our findings showed that 82.7% (5047 of 6099) of documents were in article format. Approximately all documents (99.6%) were written in English.

UT documents were published in 1926 journals, with 3403 articles (55.7%) published in 17.2% (333 journals) of all journals. FEBS Journal published the most documents of UT, followed by Biophysical Journal and Journal of Applied Polymer Science.

1991 was the only year in which the number of publications decreased in comparison with the previous year and we saw an increase in the number of publications in other years. The growth rate of UT publications per year was, approximately, 37.8%.

International collaboration among scholars is a traditional norm. Totally UT authors collaborated with scholars of 81 countries. Among them, the authors of USA with 411 co-works were in first place, followed by authors of Canada, UK, Germany and France. Probably the collaboration of UT authors with foreign peers has been affected by the English language.

As mentioned before, due to its capabilities, we used HistCite™ to draw scientific maps and depict the structure of science in UT. We made two historiographs with 60 top documents based on LCS and GCS.

We found that in the main clusters of both graphs, authors affiliated to the Chemistry Department and the Institute of Biochemistry and Biophysics were the cornerstone of science structure in UT.

Ganjali, who had published the most articles (300 articles), was present in both graphs more than others. We also found the same citation behavior in both graphs, as a result of collaboration among colleagues, self and mutual citation as well as the similarity of scientific areas.

References

1. Toffler, A.: *Powershift: Knowledge, Wealth, and Violence at the Edge of the 21st Century*. Bantam Books, New York (1990)
2. University of Tehran: *Directory of the University of Tehran (Shenasan)*. UT Publication Institute, Tehran (2008)
3. Osareh, F., Wilson, C.S.: *Collaboration in Iranian Scientific Publications*. Libri 52, 88–98 (2002)
4. Jacobs, D., Pichappan, P.: *Research Collaborations and Scientific Productivity among the Research Universities in South Africa*. In: *Proceedings International Workshop on Webometrics, Informetrics and Scientometrics & Seventh COLLNET Meeting*, pp. 433–440. INIST, Nancy (2006)

5. Lucio-Arias, D., Leydesdorff, L.: Main-path Analysis and Path-dependent Transitions in HistCiteTM-Based Historiograms. *Journal of the American Society for Information Science and Technology* 59, 1948–1962 (2008)
6. Osareh, F., McCain, K.W.: The Structure of Iranian Chemistry Research, 1990-2006: An Author Co-citation Analysis. *Journal of American Society for Information Science and Technology* 59, 2146–2155 (2008)
7. Asnafi, A., Hamidi, A., Osareh, F.: Analytical Survey and Mapping Structure of Scientific Publications in the Bibliometrics, Scientometrics, Infometrics and Webometrics Fields in Web of Science Database During 1990-2005. *Library and Information Science* 11, 161–182 (2008)

Contribution of Turkish Scholars to Earthquake Literature: The Impact of the Marmara Earthquake

Zehra Taşkın

Hacettepe University, Department of Information Management, Ankara, Turkey
ztaskin@hacettepe.edu.tr

Abstract. This paper addresses the question of whether the Marmara Earthquake of August 17, 1999, has had an impact on the contribution of Turkish scholars to the earthquake literature. We identified a total of 1,098 papers published between 1990 and 2009 by Turkish earthquake scientists. These papers were cited 7,691 times. Both the number of papers and the citations they generated increased considerably after the Marmara Earthquake. This may be explained, in part, by the increase in the number of projects being carried out since then to study the Marmara Sea basin.

Keywords: Marmara Earthquake, citation analysis, Turkish earthquake literature.

1 Introduction

Due to its geographical position, there have been many earthquakes in Turkey throughout history. The Marmara Earthquake of August 17, 1999, was one of the most destructive earthquakes in recent history. It measured 7.4 on the Richter scale and affected mainly the urban dwellers living in Istanbul and its environs [1]. According to the initial reports, the earthquake resulted in approximately 15,226 fatalities and 23,983 injured. In addition to mortalities and injuries, it caused US \$9-13 billion property damage. It was estimated that 14,444,298 inhabitants living in the Marmara region were affected by the event. The Marmara Earthquake has had negative effects on Turkish industry, education and health systems, labor force, agriculture and infrastructure [2].

This paper reviews the effects of the Marmara Earthquake on the number of publications authored by Turkish scholars. It identifies the most productive authors and institutions carrying out earthquake research in Turkey.

2 Problem Statement

The main aim of this study is to evaluate the papers on earthquakes authored by Turkish scholars between 1990 and 2009 and to determine the impact of the Marmara Earthquake of 1999. The hypothesis of this study is: "The number of publications

authored by Turkish scholars in the earthquake field increased after the Marmara Earthquake of 1999". More specifically, this study attempts to answer the following research questions:

- What types of publications are published by Turkish earthquake scientists?
- What percentage of the world's earthquake literature is generated by Turkish scholars?
- Who are the most productive authors and institutions carrying out research on earthquakes in Turkey?
- Where do Turkish earthquake scientists publish and how often are their works cited?

3 Literature Review

There have been many papers about bibliometric analysis of publications in a specific field [3], [4], [5], [6]. The earthquake literature doesn't include any work directly relevant to citation analysis. However, there are three such articles on seismic studies and earthquake engineering.

Some 1128 papers published by the Institute for Geophysics of University of Texas were evaluated in a study [7]. Papers were classified into four categories that differed significantly with respect to statistics such as lifetime citation rates, fraction of papers never cited after 10 years and cited half-life. It was concluded that reported differences in cited half-lives must be quite large to be significant.

Trifunac [8] analyzed the works of 51 academics for the aim of finding influential researchers on earthquake engineering. He used ISI's HighlyCited.com and tried to find out why there are no earthquake engineers in the category of engineering. As a result, earthquake engineering was absent in the engineering category of ISI's HighlyCited.com. Trifunac also compared female and male academics in earthquake engineering by using citation analysis methods [9].

Studies on bibliometrics and citation analysis have also been published in Turkey. They generally examined dissertations [10], [11], [12], [13].

Some 572 Turkish physics publications that appeared in Science Citation Index between 1982 and 1990 were evaluated in a study [14]. They found that papers from Turkey published in European or American journals are cited more frequently.

Papers from Hacettepe University published between 1975 and 2003 were appraised in 2004 [15]. Science Citation Index and Journal Citation Reports were used as data tools and 9688 publications were analyzed. As a conclusion, over one third of publications were published after 2000. Almost all the publications were written in English. Four, and more than four-author, publications constituted 57% of all publications. Only 8% of the publications had single authorship.

There are a few studies for the combination of earthquakes and citation analyses in print. However, Turkey hasn't yet published such an analysis. This is the first study based on citation analysis of papers on earthquakes in Turkey.

4 Methodology

ISI Web of Science has been used as a data tool for this study. Data have been collected from Science Citation Index (SCI), Social Sciences Citation Index (SSCI), Conference Proceedings Citation Index – Science (CPCI-S) and Conference Proceedings Citation Index – Social Sciences (CPCI-SS) databases.

Searches were carried out on April 27, 2010. The term “earthquake” was used as the search topic. Then, for the purpose of finding papers with addresses belonging to Turkish institutions, the terms “Turkey” or “Türkiye” were entered in the address field.

Data were analyzed by using “analyze results” and “create citation reports” options provided by the Web of Science [16]. In addition, Excel and SPSS software were used to create tables and charts.

5 Findings

5.1 Publication Count and Citations

We identified a total of 1,098 papers on earthquakes published by Turkish scholars between 1990 and 2009 [17]. These publications were cited 7,691 times and average citations per publication was 7. The average citations per publication per year was 366.24. Figure 1 shows the graph of annual publication counts for Turkey and the world.

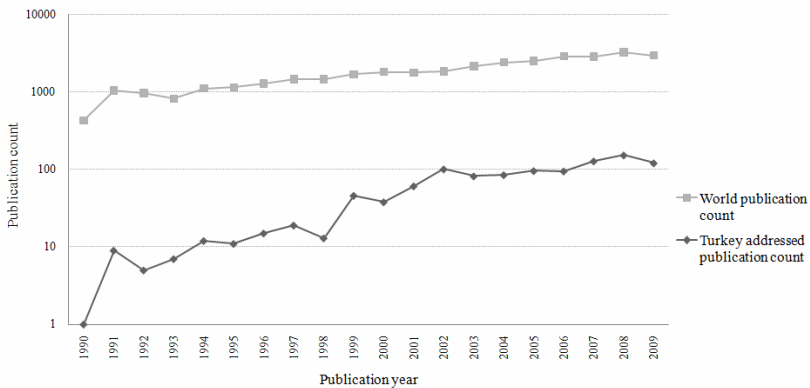


Fig. 1. The graph of publication count by years

Although the publication count had been relatively low until the end of the 1990s, there has been a huge growth since the year 2000. Commensurate with the increase in the number of publications, the number of citations to papers by Turkish scholars has also increased tremendously since the year 2000 (Figure 2). These figures clearly show that the number of papers and citations thereto have almost doubled since the Marmara Earthquake of 1999.

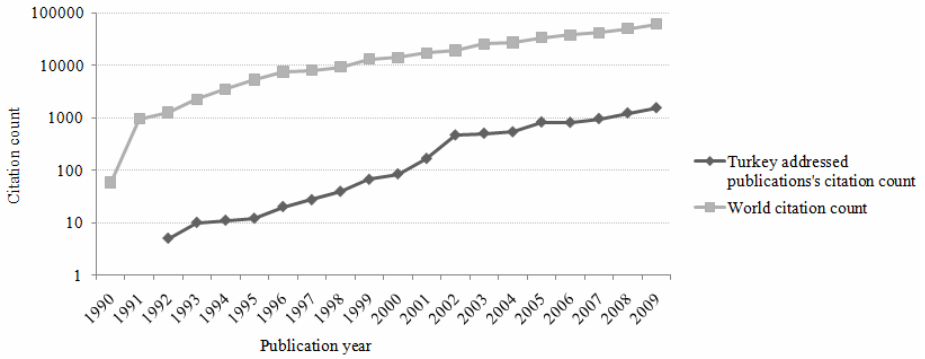


Fig. 2. The graph of citation count by years

Nearly 220 of the publications are about the Marmara Earthquake of 1999. Figure 3 clearly shows that. 55% of these publications were published between 2002 and 2005. Although almost all publications published in 2002 are about the Marmara Earthquake, the number of publications on this subject has decreased in recent years.

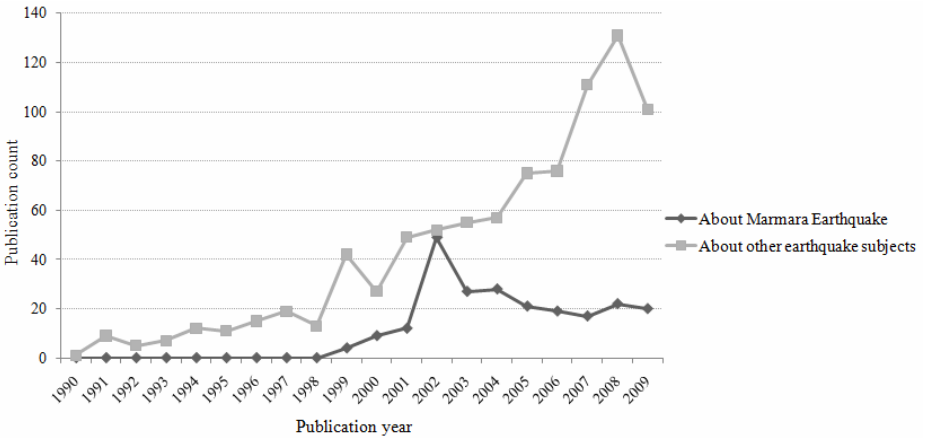


Fig. 3. Distribution of publications according to their subjects

Some 61% of publications were cited at least once. The rest have yet to be cited. Ten publications were cited more than 100 times. Figure 4 shows the distribution of publication and citation counts.

Nearly half the non-cited 431 publications were published in 2008 and 2009 (see Figure 5), which means that they have the potential of generating citations in the coming years.

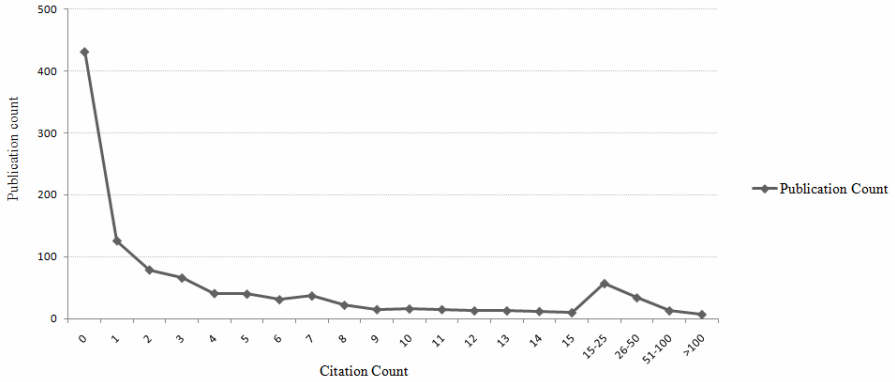


Fig. 4. Distribution of publications and citation counts

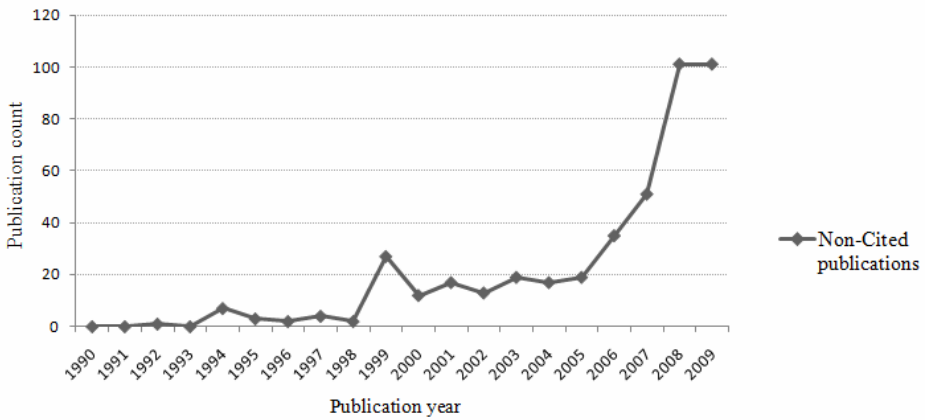


Fig. 5. Non-cited publications and their publication years

5.2 Publication Types

Almost three fourths of the publications were journal articles. In addition to these articles, there are also proceedings papers, editorial materials and letters (see Table 1).

Table 1. Document types (N=1,098)

Document Type	N	%
Article	823	74.9
Proceedings paper	215	19.5
Editorial material	26	2.3
Letter	14	1.2
Review	12	1.1
Meeting abstract	7	0.6
Note	1	0.1

5.3 Comparison with Other Countries

Some 34,721 articles were published between 1990 and 2009 in the world about earthquakes. USA ranks 1st, generating one third of the papers, while Turkey ranks 11th in terms of contribution to the earthquake literature. The Turkish contribution made up only 1% of the world's earthquake literature before 1998, whereas it quadrupled to 4% after the Marmara Earthquake of 1999.

5.4 Journals

Over 7% of all articles published by Turkish scholars appeared in one journal (Bulletin of the Seismological Society of America). The vast majority of journals published one or two articles from Turkish scholars. Journals publishing only one article constituted 60% of all journals in which contributions from Turkey appeared. Figure 6 shows the cumulative distribution of articles in accordance with journals.

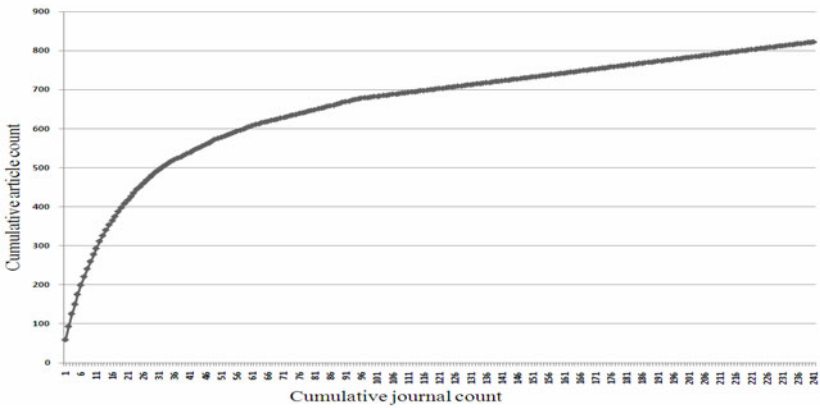


Fig. 6. Cumulative distributions of articles

Table 2. Most used journals and their impact factors [19]

Journal's Name	Number of articles	Impact Factor
Bulletin of The Seismological Society of America	60	2,199
Geophysical Journal International	33	2,219
Engineering Structures	32	1,102
Soil Dynamics And Earthquake Engineering	26	1,182
Engineering Geology	25	1,197
Structural Engineering And Mechanics	23	0,500
Earthquake Engineering & Structural Dynamics	21	1,240
Natural Hazards	21	1,142
Journal of Seismology	18	1,091
Tectonophysics	18	1,670

Some 822 articles were published in 241 journals. The impact factor is a measure of the frequency with which the “average article” in a journal has been cited in a particular year or period. Journal Citation Reports (JCR) of Thomson Reuters provides journals’ impact factors. The annual JCR impact factor is a ratio between citations and recent citable items published [18]. The most used journals and their impact factors are listed in Table 2. Turkish authors seem to have preferred international journals. They published only 12 articles in the Turkish Journal of Earth Sciences and 9 in Teknik Dergi, both journals published in Turkey.

5.5 Authors and Their Institutions

Generally, journal articles were published by authors based in big universities of Turkey. Five universities published over half the articles (see Table 3).

Table 3. Top five universities for earthquake articles

Institution Name	N	%
Bogazici University	117	14.21
Istanbul Technical University	114	13.85
Middle East Technical University	107	13.00
Istanbul University	59	7.16
Karadeniz Technical University	57	6.92

The most productive author was A. Bayraktar who published 27 articles. Dr. Aykut Barka, who died in 2002, ranked 3rd with 21 articles published between 1996 and 2002 (see Table 4).

Table 4. Most productive first ten authors

Author	N	%
Bayraktar, A	27	3.28
Erdik, M	22	2.67
Barka, A	21	2.55
Sever, MS	20	2.43
Aktar, M	19	2.30
Ergintav, S	18	2.18
Sucuoglu, H	17	2.06
Vanholder, R	17	2.06
Alptekin, O	14	1.70
Eyidogan, H	14	1.70

6 Conclusion

The number of publications authored by Turkish earthquake scientists has increased considerably since 1999. The main reason seems to be the Marmara Earthquake of

1999, as the publication and citation counts have doubled since then. The contribution of Turkey to the world's earthquake literature quadrupled as well, placing Turkey in the 11th rank among the world's countries. The hypothesis of this study is accepted.

Yet, papers by Turkish earthquake scientists have not been cited heavily in the literature. Very few papers have generated more than 100 citations although 60% of publications were cited at least once. Turkish earthquake scientists generally preferred international journals to publish their work, and 7% of their contributions appeared in a prestigious journal (*Bulletin of the Seismological Society of America*).

It is hoped that research earthquake carried out in Turkey will improve our understanding of devastating earthquakes not only in Turkey but elsewhere. If the current growth rate of publications on earthquakes continues, Turkey can be one of the top ten countries in the world conducting serious research on earthquake engineering.

Acknowledgments. I would like to thank Prof. Dr. Yaşar Tonta for his very helpful comments, suggestions and improvements on this paper.

References

1. Ozerdem, A.: Tiles, Taps and Earthquake-Proofing: Lessons for Disaster Management in Turkey. *Environment and Urbanization* 11, 177–179 (1999)
2. T.R. Prime Ministry State Planning Organization: Deprem'in Ekonomik ve Sosyal Etkileri (Economic and Social Effects of Earthquakes). Prime Ministry State Planning Organization Report, pp. 3–4 (1999)
3. Ding, Y., Chowdhury, G.G., Foo, S.S.: Journal as Markers of Intellectual Space: Journal Co-Citation Analysis of Information Retrieval Area 1987-1997. *Scientometrics* 47, 55–73 (2000)
4. Zhang, Y.: The Impact of Internet-Based Electronic Resources on Formal Scholarly Communication in the Area of Library and Information Science: a Citation Analysis. *Journal of Information Science* 24, 241–254 (1998)
5. Burrell, M.A., Hahn, T.B., Antonisse, M.J.: Understanding Information Use in a Multidisciplinary Field: A Local Citation Analysis of Neuroscience Research. *College and Research Libraries* 66, 198–210 (2005)
6. Qiu, H., Chen, Y.F.: Bibliometric Analysis of Biological Invasions Research During the Period of 1991 to 2007. *Scientometrics* 81, 601–610 (2009)
7. Frohlich, C., Resler, L.: Analysis of Publications and Citations from a Geophysics Research Institute. *Journal of the American Society for Information Science and Technology* 52, 701–713 (2001)
8. Trifunac, M.D.: On Citation Rates in Earthquake Engineering. *Soil Dynamics and Earthquake Engineering* 26, 1049–1062 (2006)
9. Trifunac, M.D.: A Note on Publication and Citation Rates of Female Academics in Earthquake Engineering. *Soil Dynamics and Earthquake Engineering* 26, 1063–1075 (2006)
10. Tonta, Y., Al, U.: Scatter and Obsolescence of Journals Cited in Theses and Dissertations of Librarianship. *Library and Information Science Research* 28, 281–296 (2006)
11. Uçak, N., Al, U.: Bilimsel İletişimin Zamana Göre Değişimi: Bir Atf Analizi Çalışması (Changes in Scholarly Communication Over Time: a Citation Analysis Study). *Bilgi Dünyası* 10, 1–22 (2009)
12. Uçak, N., Al, U.: The Differences Among Disciplines in Scholarly Communication: a Bibliometric Analysis of Theses. *Libri* 59, 166–179 (2009)

13. Bayram, Ö.: Atf Verisi ve Enformetrik Yasalar: Türk Kütüphanecilik Literatüründeki Doktora Tezleri Üzerinde Bir Uygulama (Citation Data and Informetrics Law: an Application in Doctoral Theses to the Turkish Librarianship). *Türk Kütüphaneciliği* 12, 21–32 (1998)
14. Uzun, A., Menard, A., Ozel, M.E.: Citation Status of Turkish Physics Publications in Foreign-Journals - a Global Analysis. *Scientometrics* 28, 79–87 (1993)
15. Al, P., Al, U., Bahşışođlu, H.K.: Science Citation Index'de Hacettepe Üniversitesi: 1975-2003 (Hacettepe University in the Science Citation Index: 1975-2003). *Hacettepe Üniversitesi Edebiyat Fakültesi Dergisi* 21, 229–244 (2004)
16. Thomson Reuters, <http://science.thomsonreuters.com/m/pdfs/mgr/ws-wos-8-0-0807.pdf>
17. Web of Science, <http://apps.isiknowledge.com>
18. Thomson Reuters, http://thomsonreuters.com/products_services/science/free/essays/impact_factor/
19. Journal Citation Reports, <http://apps.isiknowledge.com/JCR>

Author Index

- Akçapinar Sezer, Ebru 145
Alr, Gülten 129
- Birrell, Duncan 45
Bozkir, A. Selman 145
- da Silva, Armando Malheiro 189
Dobрева, Milena 45
- Fazeli, Mahlagha 120
Feliciati, Pierluigi 45
Foo, Schubert 95
- Gathegi, John N. 154
Güzin Mazman, S. 145
- Helvoort, A.A.J. (Jos) van 61
Horvat, Aleksandra 161
- Jivan, Naser Ezzati 120
- Kajberg, Leif 3
Keshvari, Maryam 200
- Landis, Cliff 18
Lippincott, Joan K. 1
- Mac an Airchinnigh, Mícheál 31
Majid, Shaheen 95, 110
Marsh, Corrie V. 18
- Martins, Fernanda 189
McCulloch, Emma 45
- Nagata, Haruki 74
- Oguz, Fatih 18
Osareh, Farideh 200, 211
Oğuz, Esin Sultan 3
- Ribeiro, Fernanda 189
- San, Myint Malar 110
Scepanski, Jordan M. 85
Sødring, Thomas 129
Soydal, İrem 129, 175
- Taşkın, Zehra 222
Tonta, Yaşar 175
Tun, Saw Tin Naung 110
- Ünal, Yurdağül 45
- Wells, H. Lea 85
- Yousefi, Khadije Sadat 120
- Zar, Thu 110
Zare, Amin 211
Zhang, Xue 95
Zivkovic, Daniela 161