INFuture2011

The Future of
Information Sciences

Information
Sciences and
e-Society

THE FUTURE OF INFORMATION SCIENCES

INFuture 2011

3rd International Conference
"The Future of Information Sciences:
INFuture2011 – Information Sciences and e-Society"
Zagreb, 9-11 November 2011

THE FUTURE OF
INFORMATION SCIENCES

INFUTURE2011

INFORMATION SCIENCES
AND E-SOCIETY

Edited by

Clive Billenness, Annette Hemera, Vladimir Mateljan,
Mihaela Banek Zorica, Hrvoje Stančić, Sanja Seljan

Zagreb, November 2011

CONTENTS

Preface

The third conference *The Future of Information Sciences* – **INFuture 2011: Information Sciences and e-Society**, aims to bring together researchers, professionals, businessmen and project managers from this hot topic domain. This time, it is organised by the Department of Information Sciences, Faculty of Humanities and Social Sciences, University of Zagreb under the high auspices of Croatian Academy of Sciences and Arts (HAZU).

The interaction between science, society and business applications is presented from various aspects, discussing theoretical and practical issues, such as e-government, electronic voting, mobile applications, etc. The role of e-learning in Croatian academic institutions is analyzed, as well as the e-society aspects through collaboration and in online multimedia environment. Social and media communication discuss the role of technology in society, social networks and electronic publishing. Digital preservation is considered in an archival surrounding, and as an e-science. The role of language technologies and natural language processing are analyzed through the aspects of language identification, automatic detection, web services, written and spoken word analysis and evaluation of translation tools.

INFuture2011 established cooperation with KEEP Project which aims to develop emulation services for enabling rendering of static and dynamic objects, in order to facilitate access to cultural heritage. KEEP Project will disseminate their research results at the INFuture2011 conference.

At the third, INFuture2011 conference, about 50 internationally peer-reviewed papers are to be presented, together with invited speeches and two project presentations, elaborating, through multidisciplinary approaches, the role of information sciences and e-society. About 120 participants from 12 countries will participate, together with participants from the KEEP Project.

We believe that the interaction among scientists, professionals, businessmen and investors could create a platform for building common high-quality tools and resources in the modern, e-society.

Editorial Board

E-SOCIETY

Children and e-Society: European Research Findings on Opportunities and Risks

Brian O'Neill
Dublin Institute of Technology / EU Kids Online
brian.oneill@dit.ie

Children are widely seen as central to the development of e-society: as direct beneficiaries and indeed often identified as the primary targets of information society policies, particularly those geared towards enhancing learning opportunities, access to information and building inclusiveness and participation in society. The European Union's *Digital Agenda* places a safer and better internet for children at the heart of its policy platform. And yet, more often than not, children's e-society participation has been a cause of concern and anxiety for policy makers, particularly with ever-increasing early adoption of new internet technologies and services by children and young people. Such concerns have been motivated by the responsibilities held by public agencies to ensure adequate protection for young people whilst seeking to encourage and foster children's online opportunities. Thus, e-society may be said to constitute a set of tricky policy dilemmas and challenges with regard to children's participation. To date, the balancing of risks and opportunities has been informed more by assumptions of the benefits and the dangers that e-society might pose for children and young people. *EU Kids Online*, a pan-European survey of children's use of the internet, has attempted to fill this research gap by providing the first fully comparable data on issues of risk and safety gathered directly from children themselves in 25 European countries. Research findings focus on the extent of children's embeddedness within e-society, examining dimensions of e-literacy, the availability of appropriate e-content and resilience in relation to risks encountered online. It is important to argue for greater policy and research attention to children's perspectives on e-society, as a means of fostering greater trust and participation for society as a whole.

EU Kids Online, http://www.eukidsonline.net

Tranformation of Scholarly Publishing in the Digital Era: Scholars' Point of View

Radovan Vrana
Department of Information Sciences,
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
rvrana@ffzg.hr

Summary

Scholarly publishing is one of the most important activities for scholars whose career depends on publication of their research results. To find out more about the current situation regarding scholarly publishing at the Faculty of Humanities and Social Sciences in Zagreb, Croatia, a research project was initiated. Findings showed strong orientation of scholars towards the most traditional publication channel in science - printed journal. Findings also revealed that scholars most frequently write their articles alone; sometimes they wait for more than a year for publication of their articles mainly in journals; they are seldom editors of journals, but they participate in work of editorial boards and they also participate in the peer review. The outcomes of this research showed that scholars are still communicating by using the infrastructure of the old paradigm of scientific communication based on printed information resources and that they suffer from the same problems as their colleagues around the world regarding communication of the results of their scientific research.

Key words: Scientific communication, scholarly publishing

Introduction

Scholarly publishing is an essential activity in the scientific community worldwide. It allows scholars to present and promote results of their scientific research and gain prestige for themselves and their academic institutions as well as to attract grants for new scientific research. The scientific research at universities has been communicated through the ages primarily via scholarly publishing (Honey, 2005, p. 59) or more precisely, by the help of articles printed in scientific journals, the primary vehicle for communicating and documenting results in most scientific disciplines (Lynch, 2007). Journals have a special role in science since they "(…) have a function of documenting invention and ideas and have, over time, established methods for bestowing merit and status (e.g., through peer review and recognized editorial boards)

(…)" (Lynch, 2007). The traditional scientific journal is undergoing a transformation, initiated by technological opportunities and by a series of environmental factors that will shape the future structure and functionality of publications and communication (Lougee, 2000, p. 239). This ongoing publishing revolution has generated great quantities of digital content creating opportunities for new forms of research and scholarship, qualitatively different from the traditional scholarship based on printed information resources. To find out whether the scholars are still relying on the printed information resources for publication of their research results, a research project was initiated.

Problems encountered by scholars in the process of scholarly publishing

Writing an article for publication in a scientific journal is one of the most important jobs of a scholar. Scholars write scientific articles "(…) to communicate precise information, idea, concepts in a standard format. Another goal of writing is to persuade audience on the conclusions." (Khattri, 2009, p. 187). They are very keen on publishing their articles in scientific journals of high quality. They choose carefully journals which offer the most favourable conditions for publication of their articles. Scientific journal is "(…) the main genre used by scientists to report on their work and document results of their research" (Mackenzie Owen, 2007, p. 37). Final choice of the journal is made upon scholar's perceptions about "(…) the relative qualities of the journal, the efficiency of the expected review process and the estimated likelihood of acceptance by each journal." (Miller and Harris, 2004, p. 75). Scholars are also occasionally members of the journal editorial boards where they participate in making the most important decisions about the journal publishing policies and where they encounter different problems related to communication with authors and journal editors. These problems are shared by many journals as their authors come from many world countries and make similar demands to the journals and journal editorial boards: speedy publication, fair peer review, fair communication with journal editor etc. Finally, scholars are sometimes taking part in the peer review process as reviewers. This activity is usually done in scholars' free time without getting any material compensation for the work done. Mulligan enumerated five key criteria considered by the prospective reviewers when they are making decision whether to accept or to reject this job: "(…) paper relevancy, journal reputation, the quality of the article and impact factor of the journal." (Mulligan, 2008, p. 199). In spite of the problems they encounter during their participation in the peer review process, scholars continue to accept journal articles for the review. Gisvold explained this phenomenon: for some reason scholars "(…) seem to think that being a member of the academic community carries an obligation to take part in this kind of work." (Gisvold, 2007, p. 977). Many other problems are present in the exciting world of scholarly publishing but due to the space constraints, only a selection

8

of the important problems scholars frequently encounter as authors, members of editorial boards and as reviewers have been presented in this chapter.

Research

To find out more about the experience of scholars in roles of authors, member of the journal editorial boards and peer reviewers a research was initiated. Scholars at the Faculty of Humanities and Social Sciences in Zagreb (FHSS) were chosen as a target group for this research because they are work in the biggest academic institution in the fields of humanities and social sciences in Croatia. The main hypothesis of this paper is that scholars at the FHSS still prefer traditional communication channel in science - printed journal as their principal communication channel. The purpose of this research is to identify characteristics and problems in the process of scholarly publishing in this target group of scholars. Web survey consisting of 27 closed type questions was chosen as a method of conducting the research. The research started on January 24th 2011 by sending e-mail invitations to all employees at the FHSS who regularly participate in research and teaching process at this Faculty (approximately 600 people). The research was closed on February 9th 2011 with 106 people who participated in it (response rate of 17.6 %). While the response rate may seem to be low, collected results represent a relevant snapshot of the current situation of publishing habits of scholars at the FHSS and it can serve as an orientation for future research.

Findings

The survey was divided into three parts. The first part consisted of two questions about scholars' title and field of science (social sciences or humanities). The second part aimed at collecting data about scholar's experience with publishing of their articles in scientific journals. The third part aimed at collecting data about scholar's experience as reviewers and/or journal editors or members of journal editorial boards. Due to the space constraints, only part of the survey findings will be presented in this section.

74 (71.2%) participants came from the field of humanities, and 30 participants (28.8%) came from social sciences (some participants didn't indicate the scientific field they work in). The structure of the respondents was the following: assistants (42), assistant professors (32), full professors (18), associate professors (8), senior foreign-language instructor (2), lecturer (1), professional associate (1), senior professional associate (1). The presented participant structure can be attributed to the assumption that the younger employees at the FHSS use e-mail as a preferred communication tool more frequently than their older colleagues. It must be noted that some categories, such as lecturers and foreign-language instructors, have a small number of

members, so their number will always be smaller compared to, for instance, assistants.

Q1. What type of work do you write most frequently?

Table 1: Type of work written most frequently (N=105)

	N	%
Scientific article in a journal	71	67.6
Paper in conference proceedings	25	23.8
Professional article in a journal	4	3.8
Author review	2	1.9
Some other type of published work	2	1.9
Book chapter	1	1.0

The results suggest big popularity of two types of scientific works: scientific articles and papers presented at conferences and published in the conference proceedings. They are most frequently written two types of scientific works by the participants in this survey (Table 1.). Other types of works such as professional articles in journals and author reviews are less represented. Two respondents gave additional answers: "the question is badly formatted since it doesn't allow multiple choice" and "scientific article, author review, book". The popularity of scientific articles and conference papers could be explained by their importance in the process of academic advancement at the Croatian universities. These two types of the written scientific works are also highly visible in the scientific full text databases and this could be the reason why they are written more frequently by scholars than some other types of works.

Q2. How many article manuscripts have you send to journals in the last year?

Table 2: Number of article manuscripts sent to journals in the last year (N=106)

	N	%
2-3 article manuscripts	43	40.6
One article manuscript	22	20.8
4-5 article manuscripts	22	20.8
Not a single article manuscript	7	6.6
6-7 article manuscripts	7	6.6
8-9 article manuscripts	3	2.8
10 and more article manuscripts	2	1.9

Most respondents in this research write 2-3 manuscripts a year on average (Table 2.). Equal number of scholars, approximately, writes between one and 4-5 manuscripts a year. The pressure for publication of the required number of articles necessary for the academic advancement might has lead to the increased number of manuscripts sent to scientific journals. By sending the increased number of article manuscripts to journals, authors create additional pressure on

journal editorial boards which are sometimes unable to cope with increased quantities of article manuscripts sent to them, which then results in longer periods of the peer review and longer publication time. The excessive production of article manuscripts could also lead to the decreased quality of articles written in haste and under pressure.

Q3. Number of colleagues with whom you cooperate in writing of a paper (regardless of its type)?

Table 3: Cooperation with colleagues in writing of a paper (N=105)

	N	%
I'm the only author of my papers	50	47.6
I cooperate with one colleague	30	28.6
I cooperate with 2-3 colleagues	23	21.9
I cooperate with 4-5 colleagues	1	1.0
I cooperate with 10 or more colleagues	1	1.0
I cooperate with 6-7 colleagues	0	0.0
I cooperate with 8-9 colleagues	0	0.0

Almost half of the participants in this research write their papers alone (Table 3.). Another 50.5% percent (cumulatively) cooperate with 1-3 colleagues. Cooperation is necessary especially when the paper is the result of joint effort of scholars working together on the same scientific project. The extent of the cooperation depends on the area of science. In some areas of the science like natural sciences, technology or medicine, cooperation of large number of scholars is quite common. This research confirmed the existence of moderate or limited cooperation between scholars in the fields of social sciences and humanities.

Q4. Formats of article manuscripts sent to journals?(multiple answers)

Table 4: Formats of article manuscripts sent to journals (N=102)

	N	%
In electronic format, as an attachment to the e-mail message	89	93.7
Printed on paper	23	24.2
In electronic format, on CD or DVD	15	15.8
In electronic format, by uploading article manuscript to the online journal management system	13	13.7
In electronic format, on USB memory stick	4	4.2

E-mail is by far the most popular communication channel for sending article manuscripts to the scientific journals. However, almost one quarter of the respondents still send their article manuscripts to the scientific journals in paper format and even less respondents send their article manuscripts on optical media. One should bear in mind that format requirements are set by the

scientific journals and that they are usually published in author guidelines. Communication on the internet has sped up the communication between authors and journal editors and editorial boards, and it has become preferred by many journals worldwide. It is rather surprising that only a fraction of scholars who participated in this survey uses online journal management systems for sending their article manuscripts to journals. These online systems have improved journal managing process significantly as they are decrease the number of procedures usually done by journal editors manually such as receiving article manuscript and their distribution to the reviewers and back to the authors with comments etc. By using such online systems authors are in a position to monitor the status of their article manuscripts autonomously during the whole process of article review. Usage of such online systems requires support from the IT specialists which are not always available to journals, so they use e-mail as much simpler solution for the article manuscript management instead.

Q5. According to your estimation, how long does the publication period of your article usually last?

Table 5: Duration of an article publication (N=102)

	N	%
10-12 months	29	28.4
Longer than 12 months	29	28.4
4-6 months	18	17.6
7-9 months	17	16.7
2-3 months	9	8.8
Less than one month	0	0.0

Long duration of publication can influence negatively a scholar's career. However, due to the heavy demand for publication space in scientific journals, scholars have to wait for a long period of time until their article is published even if it means waiting for 12 months or longer. More than a half of the respondents (cumulatively) in total indicated that the publication time of their articles is between 10-12 months long (Table 5.) or more than 12 months long. While this period might seem to be unusually long, sometimes it is difficult to complete the review process in short period of time in journals which have too few reviewers and journal editors overwhelmed with other duties outside journal. Such long periods for publication might influence the interest for particular articles negatively since research results presented in them might get less current in time.

Q6. Regarding the journal publication medium, in which type of journal did you publish most of your articles in the last 5 years?

Table 6: Type of journals in which scholars published most of their articles in last 5 years (N=104)

	N	%
Mostly in printed journals	92	88.5
In printed and electronic journals equally	8	7.7
Can't estimate	3	2.9
Mostly in electronic journals	1	1.0

Majority of the respondents publish their articles in printed journals (Table 6.) which indicates importance of printed journals to this group of scholars. Possible reasons for this result could be: reliability, stability and high visibility of printed journals as a communication medium in science. The popularity of printed journals suggests that electronic journals still haven't become important publishing channel for scholars who participated in this research.[1] This might be the problem for the journals existing solely in the electronic format, because they are unable to attract scholars to publish their articles in them. Scholars' orientation towards publishing in printed journals could make development of open access scientific resources in Croatia more difficult.

Q7. Regarding the geographical criterion, in which type of journal did you publish most of your articles in the last 5 years?

Table 7: Type of journals in which scholars publish their articles (N=105)

	N	%
Predominantly in journals published in Croatia	62	59.0
In Croatian and foreign journals equally	25	23.8
Predominantly in journals published outside Croatia	17	16.2
Can't estimate	1	1.0

More the a half of the scholars in this research published most of their articles in the Croatian scientific journals in the last 5 years (Table 7.) and far less (23.8%) of them published both in the Croatian scientific journals and foreign journals equally. Publishing articles predominantly in the Croatian journals might be the matter of language or targeted audience, since many of the Croatian journals in the fields of social sciences and humanities are published only in the Croatian language which makes them peripheral for the readers outside Croatia who do not understand the Croatian language. Some journals are bilingual and allow

---

[1] In this paper the term electronic journal refers to all types of on-line journals and journals distributed on the Internet and by other means of electronic communication such CD, DVD etc.

authors to send their contributions in the Croatian or in the English language, depending mostly on the type of the article, because scientific articles are usually published in English, while professional articles and author reviews are published in Croatian. The choice of the language is written in author guidelines of a particular journal. Publishing articles in languages other than the Croatian language increases visibility of the author, article and journal especially today, in the internet era.

Q8. Do you publish your articles in open access journals?

Table 8: Publication of articles in open access journals (N=105)

	N	%
No	63	60
Yes	42	40

Forty percent of the respondents publish their articles in open access journals (Table 8.). Other scholars (60%) might still be reluctant to send their articles to the open access journals because they know that the articles might not be taken into consideration for the academic advancement if there is no solid proof of quality which includes the mandatory peer review in the process of publication. Open access initiative in Croatia could benefit additionally from its promotion in the Croatian academic community to gain more recognition in the academic advancement process and to attract even more scholars to publish their articles in the open access journals.

Q9. Are you an editor in a scientific and / or a professional journal? (multiple answers)

Table 9: Being an editor in a scientific or a professional journal (N=106)

	N	%
No	98	92.5
Yes, I'm editor in one Croatian scientific / professional journal	6	5.7
Yes, I'm editor in more than one Croatian scientific / professional journal	1	0.9
Yes, I'm editor in one foreign scientific / professional journal (outside Croatia)	1	0.9
Yes, I'm editor in more than one foreign scientific / professional journal (outside Croatia)	0	0.0

Only 6 scholars are editors in one of the Croatian scientific or professional journals (Table 9.), and only 1 scholar is editors in two such journals and only 1 scholar is editor in a foreign journal. Since the number of scientific journals in social sciences and humanities in Croatia is not small, it was expected that the number of scholars acting as editors would be bigger. Job of an editor can be very demanding, and scholars are sometimes reluctant to take this job.

14

Q10. Are you a member of the editorial board of a scientific or a professional journal? (multiple answers)

Table 10: Member of the editorial board in a scientific or a professional journal (N=106)

	N	%
No	66	62.3
Yes, I'm member of editorial boards in one Croatian scientific / professional journal	25	23.6
Yes, I'm member of editorial board in more than one Croatian scientific / professional journal	10	9.4
Yes, I'm member of editorial board in one foreign scientific / professional journal (outside Croatia)	10	9.4
Yes, I'm member of editorial board in more than one foreign scientific / professional journal (outside Croatia)	2	1.9

Situation is better in case of participation of scholars in editorial boards of journals (Table 10.) where 23.6% of participants are members of the editorial board in at least one Croatian scientific journal while 9.4% are members of the editorial boards of several Croatian journals and at least one foreign journal. Large number of scholars who participate in the work of editorial boards of scientific journals is good for the journal because editorial board members can take part in the editing process and take some responsibilities from the editor. In addition diversity of a journal editorial board can contribute to the quality of topics published in the journal.

Q11. What type of problems did you encounter while reviewing a journal article manuscript as a peer reviewer (multiple answers)

Table 11: Problems (first 5 only) encountered while reviewing a journal article manuscript as a peer reviewer (multiple answers) (N=68)

	N	%
Peer review job was not paid	58	74.4
I received low quality article manuscripts	45	57.7
Peer review job took me to much of my valuable time	21	26.9
Instructions received from the journal editor on how to evaluate article manuscripts were bad	20	25.6
I receiver badly formatted article manuscripts	15	19.2

The next question was focused on problems the respondents encountered while participating in the peer review process (Table 11.). The peer review is the necessary quality control procedure and it is essential if the scientific community wants to retain quality of published scientific works. Reviewers will always encounter problems, but, they usually feel that they have an obligation

to participate in the peer review process as it is an integral part of their jobs as scholars.

Table 12: Possibility of displaying preferential treatment by the article reviewers due to author's position, title or reputation (N=104)

	N	%
Yes	5	4.8
No	63	60.6
Can't estimate	36	34.6

Peer review is one of the most important processes in science. Yet, many scholars protest when they experience slowness of the process, unfairness, preferential treatment etc. The outcome of the peer review process can determine the future of scholar's career. For that reason, scholars care greatly about the peer review process and stress that it should be freed from common problems discovered during many decades of its existence. Fairness of the peer review is very important to scholars. 60,6% of the respondents believe that they didn't experience preferential treatment during the peer review process because of their reputation, title or status which they had at the moment of article manuscript submission (Table 12.). Fairness and integrity of the peer review process is essential for the process of maintaining the quality of scientific publication.

Conclusion

Scientific communication is a very complex endeavour. To be successful it requires efforts from many parties involved in it. The current system of scientific communication is undergoing a paradigm shift. This shifting process will not happen overnight, and it requires a significant amount of effort from all parties involved in the process of scientific communication to make the shift possible. Scholars are the most important party participating in this process and as such they encounter many problems while trying to publish results of their research. To find out more about the process and problems of scholarly publishing at the FHSS, a research was carried out. The outcomes of that research showed that scholars are still communicating by using the infrastructure of the old paradigm of scientific communication based on printed information resources. Generally, authors in this research suffer from the same problems as their colleagues around the world regarding communication of the results of their scientific research. Despite the problems, the system of scientific communication and its subsystem scholarly publishing remain the most important means for dissemination of scientific knowledge. Up to now, science managed to find solutions to various problems it encountered during centuries of its existence and it is expected that it will manage to find solutions to current problems as well in order to enable future transfer of knowledge.

References

Gisvold, S.E. Challenges in scientific publishing – reflections after 14 years as Editor-in-Chief. Acta Anaesthesiologica Scandinavica 2007; 51(8): 975-978.

Honey, S.L. Preservation of electronic scholarly publishing: an analysis of three approaches. // Libraries and the Academy. 5 (2005), 1; 59-75.

Lougee,W.P. Scholarly journals in the late 20th century. // Library Collection, Acquisitions & Technical Services. 24 (2000), 2; 239-50.

Lynch, C.A. The shape of the scientific article in the developing cyberinfrastructure. // CTWatch Quarterly. 3 (2007), 3. http://www.ctwatch.org/quarterly/articles/2007/08/the-shape-of-the-scientific-article-in-the-developing-cyberinfrastructure/ (12/10/2010)

Khattri, M.B. Writing, Editing and Publishing an Article in a Scientific Journal. // Dhaulagiri Journal of Sociology and Anthropology. (2009), 3; 185-196.

Mackenzie Oweb, J. The Scientific Article in the Age of Digitization. Dordrecht : Springer, 2007.

Miller, C.T.; Harris, J.C. Scholarly Journal Publication: Conflicting Agendas for Scholars, Publishers, and Institutions. // Journal of Scholarly Publishing. 35 (2004), 2; 73-91.

Mulligan, A. Quality, certification and peer review. // Information Services & Use. 28 (2008), 3-4; 197-214.

Network of Scientific Publications in Information Science

Đilda Pečarić
Department of Information Sciences,
Faculty of Humanities and Social Science
University of Zagreb
Ivana Lučića 3, 10000 Zagreb, Croatia
dpecaric@ffzg.hr

Summary

Bibliometric methods (bibliographic coupling and co-citation) implicitly postulate impact factor and influence factor as criteria for evaluation of „knowledge maps" or „intellectual structures". However, bibliometric methods can also be used for presentations of „Historio-Bibliography", that is genealogical tree of primordial publications. Visual presentation of history of key publications which E. Garfield (2001) named historiograph. Using corpus of bibliographic data from doctoral dissertations (1978-2009) we want to show the network of nodal publications from information science in the period from 1960 until today. Historiographs for presentations of Network of scientific publications could be generated with different methods (N. P. Hummon, P. Doreian, 1989): critical path method, bibliographic coupling method, citation method, co-citation method. We use these methods in order to show, by historiograph, historical overview of key publications in information science in Croatia. Chronological presentation of the development of scientific publication network also enables chronological analysis of certain authors' roles in scientific community from researchers and scholars to predecessors.

Key words: information science, scientific publications, network

Introduction

In all bibliometric researches Lotka's law is confirmed: a small number of authors holds the largest part of scientific production. Following that logic the most cited authors have the largest influence factor. That law is also confirmed in previous research on the corpus of doctoral dissertations in information science done at Croatian universities from 1978 to 2009. Based on gathered and analysed bibliographic data about the corpus of 170 doctoral dissertations, and 28,188 analysed references, an overview of (the most) cited authors and documents in information science according to the criteria of institutions, disciplines and time periods was retrieved. Indicators of our research follow

Lotka's distribution. Analysing 13273 cited authors it is established that the biggest number of authors (76.5%) is cited only once, and only a small number of authors is frequently cited. The most cited 43 authors (0.32% of authors) is cited 1339 times, which covers 6.1% of all citations. Citation frequency of the most cited 43 authors is in range from 18 to 68 times (in average 43 authors are cited 30 times). (Đ. Pečarić, 2010)

In this paper we wish to show the network of the most cited publications in information science using analysed corpus. Furthermore, we wish to retrieve the historiograph of key authors in information science. We will use the method which was inaugurated by E. Garfield (2001), because we believe that historiograph can be a good illustration of key publications and key authors in certain periods of information science. That is, by the selection of the most cited authors and their publications presented by historigraph we can not only understand the role of certain authors and their continual contribution to information science, but we can also present cognitive networks that show the development of information science.

Methods

The results of our research will be presented in three ways. First of all, the overview of the most cited author will be shown. The overview is retrieved by author's citation analysis of 28,188 references.

The second set of data is the overview of the most cited publications in analysed corpus. The most cited publications partially overlap with the indicators of the most cited authors. The difference is not the consequence of analysis by different methods, but rather it is consequence of the fact that the most cited authors receive citations for a series of their publications, and this also makes them relevant and influential authors. However, the criteria of citation frequency of certain publication alone is not sufficient for the construction of historiograph, neither according to disciplines nor according to time periods of information science development.

For the construction of historiograph Garfield used software for the analysis of huge amount of data. His overviews (2001) use different available methods (N. P. Hummon, P. Doreian, 1989), which we simulated by the combination of hand selection of data and an overview done by software. Since we think that historiograph be in function of cognitive network overview, together with the author we also selected: a) the year of publication's first edition, not later editions (the differences between the publishing year of original publication and the publishing years of translated edition or republished edition can be huge); b) for authors that receive citations for bigger number of their publications, we used the publication year of their oldest publication; c) in historiograph overview according to periods, we also used the year in which the oldest publication of the author cited in that period was published. We believe that in this way we can obtain more precise insight in the sequence of the appearance

and real logic of influence of certain authors on the development of information science.

Overview of the most cited authors

As we said before, in all researches that were done on the presented corpus of data we find confirmation of Lotka's law: a small number of authors holds the biggest part of scientific production. Following that logic we can conclude that the most cited authors have the largest influence factor. Indicators in our research also follow Lotka's distribution. We already indicated that a small number of authors 43 (0.32%) received even 6.1% of all citations. In another publication (Đ. Pečarić, 2010) the most cited authors in information science according to the criteria of time periods, scientific disciplines, and institutions in which doctoral dissertations were done are shown.

In table 1 we present 43 most cited authors according to citation frequency over researched period.

Table 1: Overview of 43 most cited authors in information science

Author	Citation frequency	Author	Citation frequency
Plenković, M.	68	Malović, S.	26
Srića, V.	61	Horvat, A.	26
Tuđman, M.	57	Gorman, M.	25
Novosel, P.	55	Brookes, B.C.	25
Klasinc, P.P.	43	Zelenika, R.	25
Strànský, Z.Z.	42	Prelog, N.	24
Verona, E.	41	Bauer, A.	23
Maroević, I.	40	Line, M.B.	22
Žiljak, V.	39	Mihajlov, A.I.	22
Martin, J.	38	Anić, V.	21
Lasić-Lazić, J.	37	Kržak, M.	21
Aparac, T.	36	Borgman, C.L.	21
Shera, J.H.	35	Lancaster, F.W.	21
Garfield, E.	34	Burrell, Q.L.	20
Saračević, T.	33	Price, D.J. de S.	19
Vreg, F.	32	Topolovec, V.	19
Lubetzky, S.	31	Belkin, N.J.	19
Žugaj, M.	30	Castells, M.	19
Brumec, J.	29	Buckland, M.K.	18
Šola, T.	27	Đorđević, J.	18

Author	Citation frequency		Author	Citation frequency
Foskett, D.J.	18		Žontar, J.	18
Petz, B.	18			

Data in table 1 can serve for the comparison with similar, or related researches. Unfortunately, there are only a few of them that pursue bibliometric analysis of information science in Croatia. Likewise, researches in other scientific environments are neither methodologically nor in time correlation with our research. However, although we cannot compare those research data wholly with our research, we can find mutual points of overlapping (M. Tuđman, et al., 1984 and 1988; H. D. White and B. C. Griffith, 1981; H. D. White and K. W. McCain, 1998). Mutual points of overlapping are obviously the most cited authors in different scientific communities that are research subjects in already mentioned researches. If we compare the most cited authors in the research M. Tuđman, et. al. (1988), which pursues cited analysis of 374 master theses done at CSLDIS[1] from 1961 to 1984, then among most cited authors in both researches are:

> *D. J. Foskett, E. Garfield, L. W. Lancaster, J. Martin, A. I. Mihajlov, V. Srića, J. H. Shera, E. Verona.*

The comparison of the most cited authors also partially made possible by the research of H. D. White and K. W. McCain (1998). This research analyses 120 most cited authors in 12 leading information science journals in English speaking areas. The overlapping of the most cited authors between our research and the research of H. D. White and K. W. McCain is the following:

> *N. J. Belkin, C. L. Borgman, B. C. Brookes, E. Garfield, M. Gorman, F. W. Lancaster, M. B. Line, J. Martin, D.J. de S. Price, T. Saračević.*

Also, overlapping among all three researches exists (M. Tuđman, et. al.; H. D. White i K. W. McCain; Đ. Pečarić):

> *E. Garfield, F. W. Lancaster, J. Martin.*

The comparison of most cited authors in different researches do not allow us to draw far-reaching conclusions. As we already point out, one reason is that researches include different scientific communities in different time periods, and the second reason is that researchers use different methodologies. However, from these indications it can be concluded that core authors mutual to all scientific communities i.e. information sciences communities exist, although those authors i.e. core members, change over a time. We can assume that an alteration in the sequence and frequency of authors' citation comes together

---

[1] Centre for the Study of Librarianship, Documentation and Information Sciences, University of Zagreb, Croatia.

with the alterations in research topics and interests in certain scientific communities.

Overview of the most cited publications according to disciplines

After we gained insight into the most cited authors, we can raise the question which publications received the most citations? By answering that question we can also obtain the answer to the question which authors are cohesive factors in information science, but also which publications homogenised the entire area of information science or certain disciplines.

For that reason, according to citation frequency of certain publications criteria, we extracted the most cited publications, not authors (previous tables were done based on the authors' citation frequency and not the frequency of their publications). The overview of the most cited publications according to disciplines, with marked criterion of selection in the brackets, follows.

Archival and documentation science (citation frequency bigger than 2):
- **Beuc, I.** Arhivistika.
- **Antoljak, S.** Pomoćne istorijske nauke.
- **Kolanović, J.** Arhivistika i povijest upravnih institucija
- **Tuđman, M.** Teorija informacijske znanosti.

Information systems (citation frequency bigger than 5):
- **Srića, V.** Uvod u sistemski inženjering.
- **Strahonja, V.; Varga, M.; Pavlić, M.** Projektiranje informacijskih sustava.
- **Srića, V.** Sistem, informacija, kompjutor.
- **Lazarević, B.; Jovanović, V.; Vučković, M.** Projektovanje informacijskih sistema.
- **Radovan, M.** Projektiranje informacijskih sistema.
- **Tkalac, S.** Relacijski model podataka.

Information science (citation frequency bigger than 3):
- **Anić, V.** Rječnik hrvatskoga jezika
- **Boras, D.** Teorija i pravila segmentacije teksta na hrvatskom jeziku.
- **Chomsky, N.** Aspects of the Theory of Syntax
- **Kržak, M.** Serbo-Croatian Morpho-Spelling.

Librarianship (citation frequency bigger than 4):
- **Verona, E.** Pravilnik i priručnik za izradbu abecednih kataloga.
- **Verona, E.** Abecedni katalog u teoriji i praksi.
- **Aparac, T.** Teorijske osnove knjižnične znanosti
- **Beck, H.** Klassifikation und Informationswiedergewinnung.

Communicology (citation frequency bigger than 5):
- **Vreg, F.** Društveno komuniciranje.

- **Eco, U.** Kultura, informacija, komunikacija.
- **Novosel, P.** Delegatsko informiranje.
- **Plenković, M.** Teorija i praksa javnog komuniciranja.
- **Supek, R.** Ispitivanje javnog mnijenja.
- **Plenković, M.** Demokratizacija masmedija.
- **Plenković, M.** Suvremena radiotelevizijska retorika.
- **Plenković, M.** Teorija i praksa javnog komuniciranja.
- **Supek, R.** Ispitivanje javnog mnijenja.

Museology (citation frequency bigger than 3):

- **Maroević, I.** Uvod u muzeologiju.
- **Strànský, Z. Z.** Pojam muzeologije.
- **Strànský, Z. Z.** Temelji opće muzeologije.

From previous data we can notice several indicators. Firstly, relatively small number of publications is cited frequently. Secondly, when we look at the titles we can see that the most cited publications are ones that give theoretical frames, that is "general theory" of disciplines, thematic areas, and also norms, regulations and guides for certain professions. Thirdly, the most cited publications are mostly publications done by domestic authors, that is, publications published in Croatian (Z. Z. Strànský, F. Vreg, U. Eco). Fourthly, some authors of the most cited publications are not among the most cited authors (e.g. S. Antoljak). Fifth, some of the most cited authors are not the authors of the most cited publications. Citation frequency of certain authors is large because larger number of their publications receive citations. That is why citation frequency of publications for some of the most cited authors is small. Just to illustrate that point, following authors are the outmost case of the authors that are among most cited authors, but not one of their publications is cited more than once[2]:

Archival science: P. P. Klasinc (26); M. Kovačević (12); M. Milošević (9); M. Modrušan (9); M. Novak (10); E. Pusić (10); V. Žiljak (10).

Information science: D. W. Allen (9); A. Bookstein (8); B. C. Brookes (16); Q. L. Burrell (19); Y. S. Chen (11); D. R. Cruickshank (10); L. Egghe (13).

Information systems: S. Dobrenić (13); P. P. Klasinc (17).

Communicology: V. Lamza (17).

Librarianship: C. L. Borgman (16).

Museology: J. Beneš (15).

---

[2] In the brackets, behind authors' name is indicator of citation frequency in certain information science disciplines.

Points four and five alert us that citation frequency as the only indicator cannot be sufficient for understanding impact factor. In other words, the data about most cited authors per se are insufficient for understanding influence and/or impact of those authors in both certain disciplines and in the entire area of information science.

Historiograph or overview of cognitive networks

Graph 1 shows the overview of the most cited authors in information science. And graphs 2 to 4 show the overview of the most cited authors in doctoral dissertations in information science done at the universities in Croatia from 1978 to 2009, according to periods. These tables differ from the overview of the most cited authors which we retrieved by the citation method, that is co-citation method on the same corpus of data (Đ. Pečarić, 2009, Đ. Pečarić, M. Tuđman, 2009). The clusters of the most cited authors retrieved by co-citation analysis group authors according to citation frequency and mutual relationships. By co-citation method we can discern poles made of key authors in certain cluster, from the authors with a small number of mutual relationships that are positioned on the cluster periphery.

Although, we used the same set of data as in co-citation analysis, by historiographs (graphs 1 to 4) allow us to obtain chronological overview of the most cited authors. Historigraph is the result of the selection of the most cited authors by citation method, but the data are shown in the time sequence of the publishing of scientific publications. We have already mentioned that certain correction of data was done. It was done in the following manner: the position of certain authors on time axis is defined according to the year of first edition of cited publication (not by the year of cited reprinted editions, republished editions or the year of cited translated edition). Using this method, historiograph allowed us to obtain an overview of the most cited authors in broader context: it points to the source year of the published publication, and it also shows and connects the most cited authors in the time sequence. In this overview citation frequency is not recognisable, but it is implicitly present on graph (by the selection of the most cited authors). Also co-citation relationships between certain authors disappear.

Graph 1: Historiograph – 43 most cited authors in information science from 1978 to 2009.
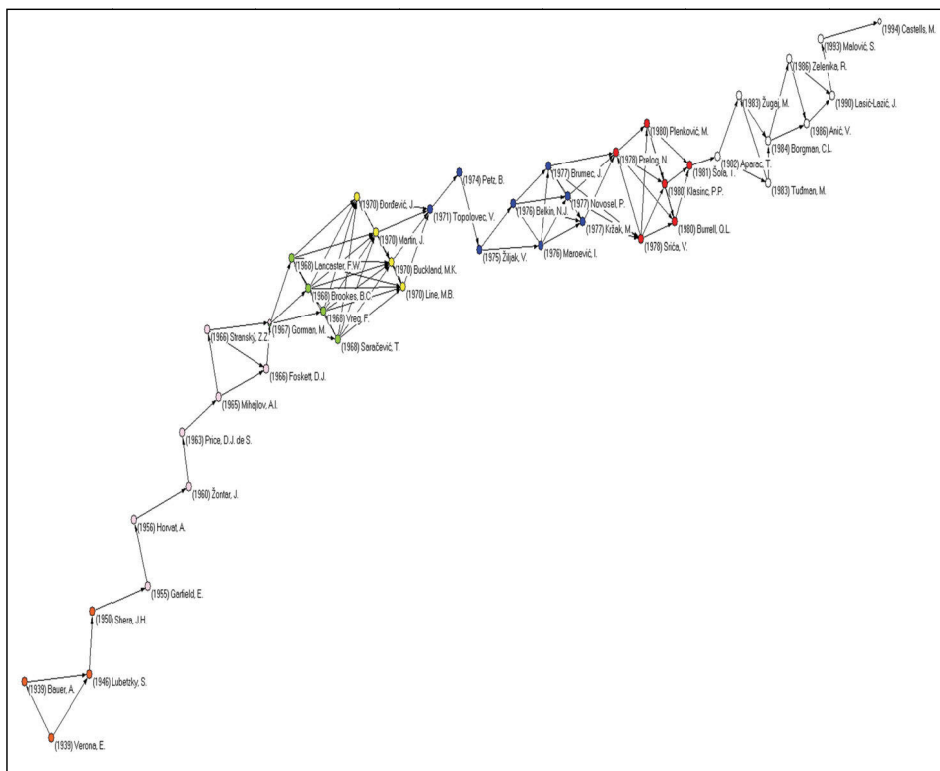
Graph 2: Historiograph – The most cited authors in information science in time period from 1978 to 1989.
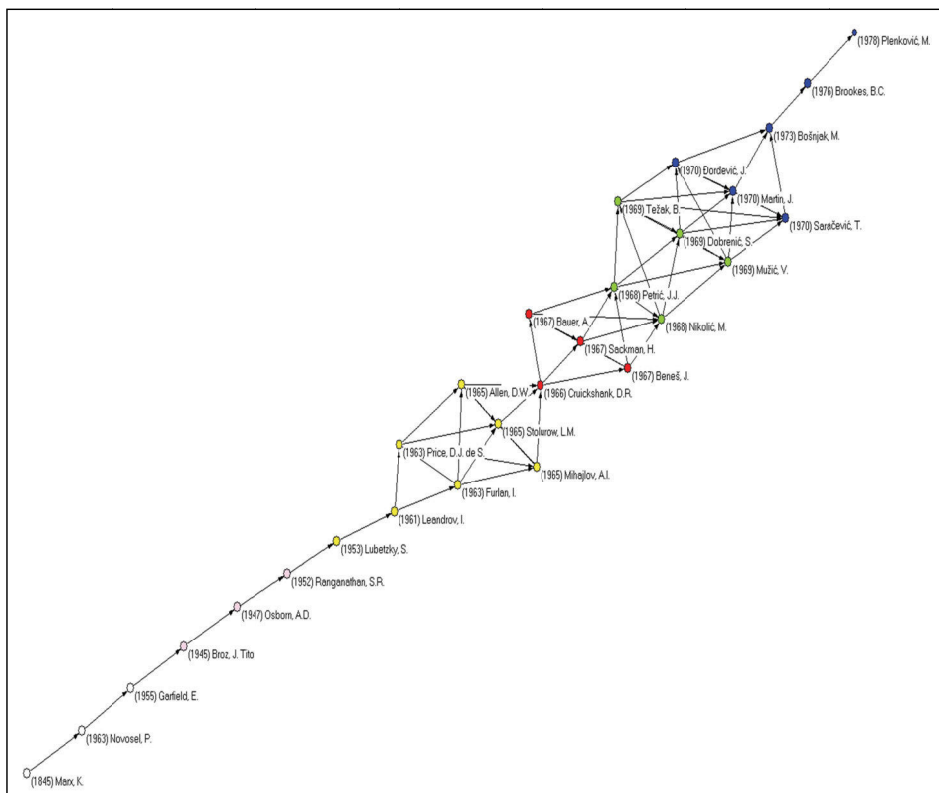
Graph 3: Historiograph – The most cited authors in information science in time period from 1990 to 1999.
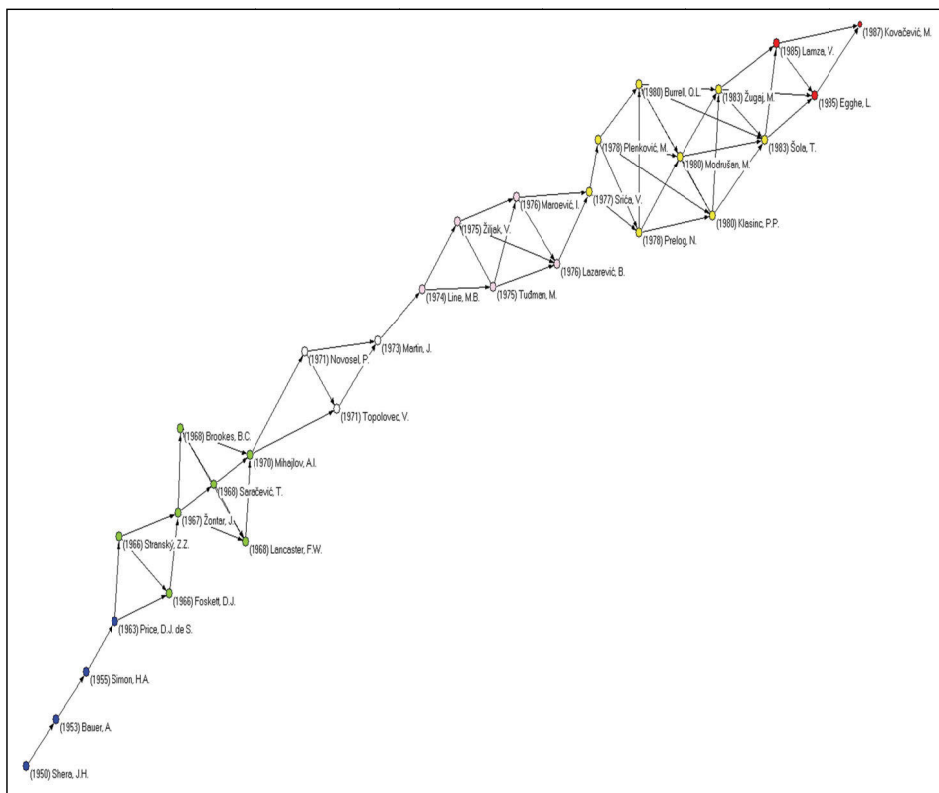
Graph 4: Historiograph – The most cited authors in information science in time period from 2000 to 2009.
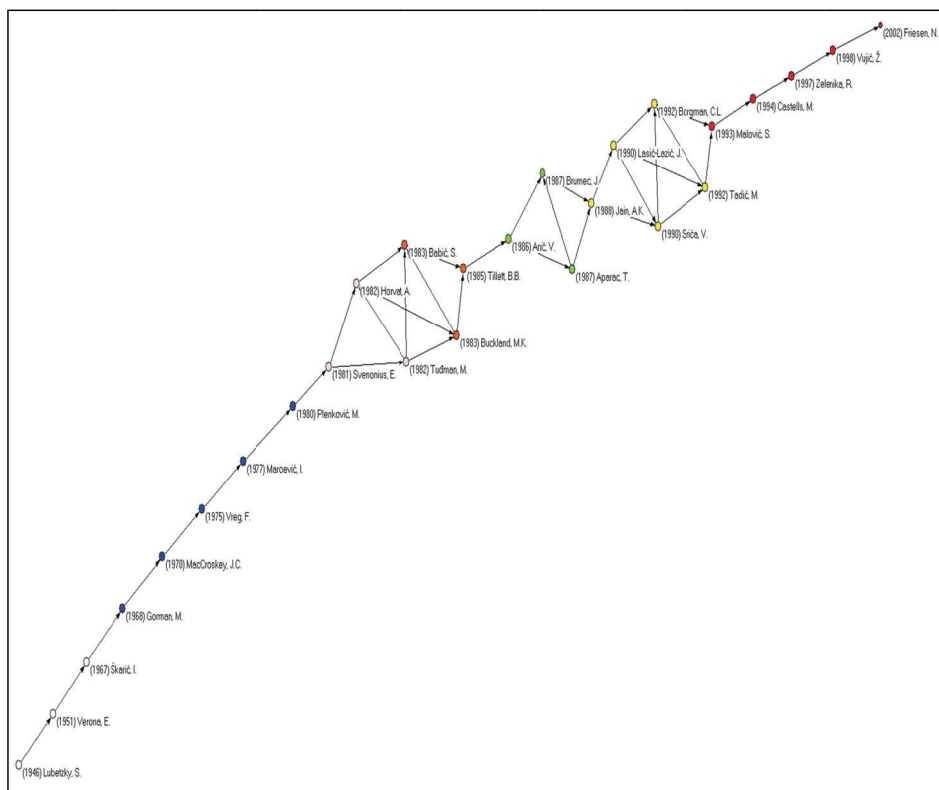
However, chronological links between authors are established, so we get an insight into the time of the appearance of certain authors and the life span of their influence on scientific community. That is why histograph can be used as an indicator of cognitive network (graph 1), and cognitive networks (graphs 2 to 4) if we follow authors' citation dynamics according to periods. In other words, histograph is the answer to the question who the most cited authors in certain scientific disciplines over time are. In combination with the previous overview of the most cited publications, historiograph can be used for the overview of the holders of the paradigm of certain disciplines, or key authors for some specific scientific area or topic.

As we cannot give detailed analysis of shown graphs here, we will point out several main characteristics.

The ratio of foreign and domestic authors on all graphs is similar (graph 1: from 43 authors 23 are foreign authors; graph 2: from 28 authors 15 are foreign;

graph 3: from 30 authors 13 are foreign; graph 4: from 27 authors 10 are foreign).

On graph 1 (shows historiograph for 43 most cited authors in the entire researched period) we can notice the sequence of authors that precede the time of establishment of information science. The most cited publication of domestic authors (A. Bauer and E. Verona) were done as early as 1939, and "global" predecessors H. Shera, E. Garfield, D. de Solla Price and A. I. Mihajlov are present in this corpus of data as well.

More authors published their first cited publication from 1968 to 1970, and again from 1976 to 1981. In the period from 1968 to 1970 foreign authors that published first cited publication are: T. Saračević, F. W. Lancaster, B. C. Brooks, F. Vreg, M. B. Line, M. K. Buckland, J. Martin, J. Đorđević. In the period from 1976 to 1981 published publications of domestic authors are mostly among the most cited publications: N. J. Belkin, I. Maroević, J. Brumec, P. Novosel, M. Kržak, V. Srića, N. Prelog, Q. L. Burrel, P. P. Klasinc, M. Plenković, T. Šola.

It is interesting that on all graphs all cited publications are older than the half-life of cited literature; half-life of analysed 28,188 references is 7.5 years. Citation half-life it remains 7.5 years for the analysed references. This confirms our postulate about the possibility of recognising the role of scientists as predecessors, scholars and researchers by bibliometric methods, and thus also by historigraps (Đ. Pečarić, M. Tuđman, 2009).

On graph 2 (the first period in the research of the development of information science: 1978-1989) among predecessors are both K. Marx and J. Broz Tito. This is yet another indicator that the ruling ideology took its toll on scientific production in one-party, undemocratic systems. The largest number of the most cited authors in this period published their publications from 1963 to 1970.

It is interesting that on graph 3 (the overview of the most cited authors in information science in period from 1990 to 1999) in the lower part of the graph we can notice the dominance of foreign authors (their publications were published from 1955 to1974), and in the upper part of the graph dominant authors are domestic authors (their publications were published from 1975 to 1987)

These differences are not noticeable in the last observed period (graph 4), although only 10 foreign authors are present from overall 27 most cited authors. However, what can be noticed here is that the life span of the most cited authors is large: from 1946 to 2002.

If we would compare data in all four historiographs we could notice several other phenomena. First of all, we can notice the differences in the lists of the most cited authors according to periods. Some authors are repeated and some authors, in time, cease to be cited frequently. But the second indicator is also interesting. Authors that receive citations in different periods frequently receive citations for other publications in new period. This indicates double dynamic of

development: scientific areas develop and in the course of time new authors with new ideas and publications enter But at the same time authors themselves develop, because their publications cause interest of new researchers, and old publications – publications that were among the most cited publications in previous periods – are not cited any more.

Conclusion

The method of historiograph was illustrated by E. Garfield in 2001, however until today it has not found broader implementation, particularly not in Croatia.
In this paper we advocate the use of historiograph for the overview of scientific development, more precisely for the development of information science. We believe that this technique of historiography provides far better diachronical overview of the development of certain scientific areas than would be possible to show by clusters, both the clusters of co-authors and clusters of co-words.
There certainly exists a number logics and techniques to obtain historiographs. In this paper we presented only one of possible approaches. Our aim was not to reconstruct scientific paradigm of information science, i.e. key authors, as much as to advocate a new method and technique of research and presentation of the development of information science.
A particular challenge would be the development of appropriate software for the preparation of historigraph. It should be made in such a manner as to enable a larger number of researchers to use this method. Alike challenge is the development of historiograph by key words. This technique would improve the research of the development of scientific disciplines. We believe that this technique would enable scientists to follow topics and thematic areas over longer time periods.

References

Garfield, Eugene. From Bibliographic Coupling to Co-Citation Analysis via Algorithmic Historio_Bibliography. 27.11.2001. URL: http://EugeneGarfield.org (17.023.2011)

Hummon, N.O.; Doreian, P. Connectivity in a Citation Network: The Development of DNA Theory. ISocial Networks 11 (1989),1, 39-63

Pečarić, Đ. Relationship between scientific paradigm and research front. On Example of Information Science Research Production. // INFuture2009 – Digital Resources and Knowledge Sharing / Stančić, H.; Seljan, S.; Bawden, D.; Lasić-Lazić, J.; Slavić, A. (ed.). Zagreb: Department of Information Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, 2009, 767-777.

Pečarić, Đilda. Razvoj informacijskih znanosti u Hrvatskoj. Bibliometrijska analiza doktorskih disertacija iz informacijskih znanosti 1978.-2007. (Doctoral dissertation). Zagreb : Filozofski fakultet Sveučilišta u Zagrebu, 2010.

Pečarić, Đ., Tuđman, M. Predecessors, Scholars and Researchers in Information Sciences. Contribution to Methodology for Bibliometrics Anaysis of Scientific Paradigms. // INFuture2009 – Digital Resources and Knowledge Sharing / Stančić, H.; Seljan, S.; Bawden, D.; Lasić-Lazić, J.; Slavić, A. (ed.). Zagreb: Department of Information Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, 2009, 713-726.

Pečarić, Đ., Tuđman, M. Theoretical difference between impact factor and influence factor. //*CECIIS 2009 Proceedings*/ Aurer, B.; Bača, M.; Rabuzin, K. (ed.). Varaždin: Faculty of Organization and Informatics, 2009, 23-28.

Tuđman, M., Milas, M., Tudor-Šilović, N., Boras, D. Bibliometric Analysis of Master Theses in Information Sciences (Postgraduate Studies in Librarianship, Archivistics, Museology and Information Science, 1961-1984, Zagreb, Yugoslavia). // *Informatologia Yugoslavica* 6 (1984), 31-40.

Tuđman, M., Tudor-Šilović, N., Boras, D. Milas-Bracović, M. A Literature Measure of Scientific Communication: Co-Citation Analysis of Master Theses in Information Sciences in Yugoslavia, 1961-1984. // *Information Research. Research methods in library and information science/* Tudor-Šilović, N.; Mihel, I. (ed.). London : Taylor Graham, 1988, 225-247.

White, H.D.; Griffith, B.C. Author Cocitaion: A Literature Measure of Intelectual Structure. // *Journal of ASIS.* 32 (1981), 2, 163-171.

White, Howard D.; McCain, Katherine W. Visualizing a discipline: An Author Co-Citation Analysis of Information Science, 1972-1995. // *Journal of the American Society for Information Science.* 49 (1998), 4, 327-355.

Usage of Open Access Institutional Repositories in Some of the European Peripheral Scientific Communities

Ivana Hebrang Grgić

Faculty of Humanities and Social Sciences University of Zagreb
I. Lučića 3, Zagreb, Croatia
ivana.grgic@ffzg.hr

Summary

In the introduction, institutional repositories and Open Access movement will be explained. In literature review, citation impact of Open Access articles will be discussed based on several researches (e.g. by Harnad, Brody, Kennicutt, Kurtz etc.). Further on, differences between scientifically mainstream and peripheral countries will be emphasized. Two important factors in defining peripheral scientific communities – language and economy – will be explained. Scientists in peripheral scientific communities sometimes communicate in "small" language (although the languages usually have long traditions, they are spoken and understood by relatively small number of scientists), and sometimes their countries' economy is marked as "developing" or "semi-developed". We will not discuss scientific communication problems in developing countries, but the problems of seven European semi-developed and developed countries with official languages other than English (Croatia, Slovenia, Bulgaria, Hungary, Italy, Poland and Greece). Institution repositories, especially if they are Open Access repositories, can significantly increase research impact of their institutions. Number and productivity of Open Access institutional repositories in the seven scientifically peripheral countries will be discussed.

Key words: Open Access, institutional repository, research impact, scientifically peripheral country

Introduction

Open Access institutional repositories (OA IRs) can play an important role in improving visibility and impact of research output of any scientific institution. Thanks to such repositories, scientific research results can be accessed anywhere in the world. As the unofficial language of scientific communication is English, some scientists who work in countries with official languages other then English, may have visibility problems when they publish scientific articles written in their mother tongue. They can also have problems if they publish

33

their articles in English in their local scientific journals that are not accessible to wider scientific community. OA IRs can increase visibility and impact of research results published by the scientists in scientifically peripheral countries.

Literature review

As defined by Lynch in 2003, institutional repository is a set of services that a scientific institution offers to the members of its community for management of digital materials produced by employees (or students) of the institution. It facilitates access to the materials, their distribution, long-term preservation and organization.[1]

According to *SPARC* (*Scholarly Publishing and Academic Resources Coalition*) definition, institutional repositories are digital collections capturing and preserving the intellectual output of a single university or a multiple institution community of colleges and universities. An OA IR has to be institutionally defined, scholarly, cumulative, perpetual, open and interoperable.[2]

Access to materials in an institutional repository can be free for anyone using Internet, or restricted to some users (e. g. registered users, employees and/or students of the institution, users who paid for the access). We will here discus Open Access repositories - repositories with no access barriers.

Open Access movement was firstly defined in the *Budapest Open Access Initiative* (*BOAI*) in 2001 as free and unrestricted online availability of scientific journal literature. Open Access should permit any users "to read, download, copy, distribute... the full texts of the articles, and use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the Internet itself."[3] The *BOAI* recommends two ways to attain Open Access to scientific literature. The first one is self-archiving, i. e. depositing full-texts of scientific papers (that will later be called Open Access repositories). The second way is publishing articles in new generation journals - Open Access journals. OA repositories are sometimes called "green route to OA" or "green road" because publishers have to give "green light" for self-archiving articles published in their journals.[4]

---

[1] Lynch, Clifford A. Institutional repositories : essential infrastructure for scholarship in the digital age. // Libraries and the Academy, 2, 3(2003), pp. 327-336.

[2] Association of Research Libraries. The case for institutional repositories : a SPARC position paper. 2002. http://www.arl.org/sparc/bm~doc/ir_final_release_102.pdf (accessed May 02, 2011).

[3] Budapest Open Access Initiative. http://www.soros.org/openaccess/read.shtml (accessed May 02, 2011).

[4] Harnad, Stevan et al. The access/impact problem and the green and gold roads to Open Access. // Serials Review 30, no. 4 (2004), available at: http://eprints.ecs.soton.ac.uk/10209/1/impact.html (accessed May 02, 2011).

In the early 2000s OA proponents had to prove the impact and citation advantage of OA articles. Without proving it, scientists would have never totally accepted OA as a way of communicating in science. According to T. Brody and S. Harnad, OA increases number of potential article users by adding the users who would otherwise have been unable to access it because of price barriers. Access is not a sufficient precondition for citation, but it is a necessary one.[5] G. J. Schwartz and R. C. Kennicutt proved in 2004 that the articles self-archived as pre-prints are cited earlier than those published and/or archived as post-prints. Also, the articles from the first group are twice as cited as the articles from the second group. The reason is the fact that an average article from the first group is accessible 12 months earlier than conventionally published articles.[6] Other authors to prove positive correlation between OA and the number of citations were Shin in 2003,[7] Moed in 2005,[8] H. Besemer in 2006,[9] Eysenbach in 2006[10] and others.

In this paper we will base our discussion on the fact that the OA advantage is proved and based on open access, early access, quality advantage and usage advantage.[11]

Defining scientifically peripheral countries

Two factors are important when distinguishing between scientifically mainstream and scientifically peripheral countries - language and economy. Global scientific community communicates in English through high-quality scientific journals, and all the countries whose scientists do not communicate in English

[5] Harnad, Stevan; Brody, Tim. Comparing the impact of Open Access (OA) vs. non-OA articles in the same journals. // D-lib Magazine 10, 6(2004), available at:
http://www.dlib.org/dlib/june04/harnad/06harnad.html (accessed May 02, 2011).

[6] Schwartz, Greg J.; Kennicutt, Robert C. Demographic and citation trends in Astrophysical Journal papers and preprints. 2004. http://www.publingresearch.net/Citations-SummaryPaper 3_000.pdf (accessed May 05, 2011).

[7] Shin, Eun-Ja. Do impact factors change with a change of medium: a comparison of impact factors when publication is by paper and through parallel publishing. // Journal of information science 29, 6(2003), pp. 527-533.

[8] Moed, Henk F. Statistical relationships between downloads and citations at the level of individual documents within a single journal. // Journal of the American Society for Information Science and Technology 56, 10(2005), pp. 1088-1097.

[9] Besemer, Hugo. Gathering evidence about the effectiveness of "Open Access" publishing policies in agriculture. 2006. http://agriscontent.wordpress.com (accessed May 05, 2011).

[10] Eysenbach, Gunther. Citation advantage of Open Access articles. PloS Biology. 2006. http://biology.plosjournals.org/archive/1545-7885/4/5/pdf/10.1371_journal.pbio.0040157-S.pdf, (accessed May, 10, 2011).

[11] Harnad, Stevan; Hajjem, Chawki. The Open Access citation advantage: quality advantage or quality bias? 2007. http://eprints.ecs.soton.ac.uk/13328/1/moed.html (accessed May 05, 2011).

may be considered scientifically peripheral. Official languages of countries they live in can have a long tradition, but are not the usual channel for scientific communication. That is the reason why the countries are considered scientifically peripheral. Another important factor is economy. Some of scientifically peripheral countries that will be discussed later are highly developed, but some of them are countries with transition economies. A transition economy is an economy which is changing from a centrally planned economy to a free market. The process has been applied in some Central European countries (Bulgaria, Croatia, Bosnia and Herzegovina…).[12]

Scientists in scientifically peripheral countries have two serious access problems. The first one is the same as the problem that scientists in scientifically mainstream countries have – price barriers that block access to high quality scientific journals. Another problem is visibility of their own research results in global scientific community. Scientists from those countries occasionally publish research results in international high-quality journals, but most frequently they publish them in local journals that do not reach global scientific community (usually because of language barriers that minimize international subscriptions, readership and influence).[13] Contemporary scientific communication system enhances the differences between journals from developed and semi-developed (or developing) countries.[14] Research results from peripheral scientific communities are invisible to the wider scientific community – there is a danger for those research results to become a part of so called "lost science". Some scientifically peripheral countries, such as Croatia, have one more characteristic – Croatian scientific journal publishers are mainly not-for-profit organisations (universities, faculties, departments, associations…). Croatian *Ministry of Science, Education and Sports* (*MSES*) gives grants only to not-for-profit scientific publishers. Another *MSES'*s funding criteria gives priority to the journals that have at least some content (not necessarily full-text articles) available on the Internet.[15] Some European scientifically peripheral countries' experiences with OA institutional repositories will be described. Their practice will be compared to Croatian practice and some recommendations for the future will be emphasized.

---

[12] International Monetary Fund. Transition economies : an IMF perspective on progress and prospects. http://www.imf.org/external/np/exr/ib/2000/110300.htm#I (accessed May 05, 2011).

[13] Stojanovski, Jadranka; Petrak, Jelka; Macan, Bojan. The Croatian national Open Access journal platform. // Learned Publishing 22, 4(2009), pp. 263-273.

[14] Marušić, Ana; Marušić, Matko. Small scientific journals from small countries: braking from a vicious circle of inadequacy. // Croatian Medical journal 40, 4(1999), pp. 508-514.

[15] Hebrang Grgić, Ivana. Open Access to scientific information in Croatia : increasing research impact of a scientifically peripheral country. Saarbrücken : Lambert Academic Publishing, 2011, p 54.

OA IRs in Croatia and some other European countries

The starting point for our analysis of OA IRs in seven European scientifically peripheral countries will be two databases – *Directory of Open Access Repositories* (*DOAR*)[16]; *Registry of Open Access Repositories* (*ROAR*)[17] – and the report that resulted from a 2010 meeting *Open Access in South European Countries*.[18] The databases and the repositories were searched at the end of April 2011. Although some aspects of institutional repositories are not comparable, we will show the total number of repositories, number or IRs, language of deposited items, material types and software used in IRs in seven European countries – Croatia, Italy, Hungary, Poland, Slovenia, Bulgaria and Greece.

According to the *DOAR*, there were five OA repositories (4 of them IRs) in Croatia in April 2011 – *Digital Archive of the Faculty of Humanities and Social Sciences* at the *University of Zagreb*, *Repository of the Faculty of Mechanical engineering and Naval Architecture* at the *University of Zagreb* (*FAMENA PhD Collection*), *University of Zagreb Medical School Repository*, *FOI* (*Faculty of Organization and Informatics*) *digital library* and *Hrčak* portal. According the *ROAR*, there were, in April 2011, three repositories – all of them mentioned earlier –*University of Zagreb Medical School Repository*, *Digital Archive of the Faculty of Humanities and Social Sciences* at the *University of Zagreb* and *Hrčak* portal. Analysing software used in Croatian repositories, we can see that two of them use EPrints, one uses dSpace and two repositories use some other software. All the repositories have items written in Croatian and three of them also have items written in English. Material types in Croatian OA IRs are journal articles (in 3 repositories), theses (in 3 repositories), books or chapters (in two repositories), conference papers (in two repositories) and there are also other material types (presentations, working papers, unpublished material…).

If we analyse content of and access to all the repositories, we can see that not all of them meet the full definition of an OA IR. The *FOI* digital library has an intention to become an OA IR, but is in a test phase and has only 5 items, all of them administrative documents of the institution.[19] The *Hrčak* portal is a valuable project but it is not really an OA repository – it is a platform for access to Croatian OA journals. The *Digital Archive of the Faculty of Humanities and Social Sciences* at the *University of Zagreb* (720 items) has an intention to be an OA IR, it operates on the EPrints software and is OAI PMH compliant but it

---

[16] DOAR: Directory of Open Access Repositories. www.opendoar.org (accessed May 11, 2011).

[17] ROAR: Registry of Open Access Repositories. 2010. http://roar.eprints.org (accessed May 11, 2011).

[18] Open Access in Southern European Countries.
http://www.accesoabierto.net/sites/accesoabierto.net/files/OASouthEurope.pdf (accessed May 11, 2011).

[19] FOI dlib. http://dlib.foi.hr/handle/10439/1 (accessed May 11, 2011).

still does not allow OA to all of its content (to access some full-texts registration is required). The repository archives mainly diploma thesis, although it has an intention to archive other material types. *Repository of the Faculty of Mechanical engineering and Naval Architecture* at the *University of Zagreb* (684 items) does not allow OA to all of the deposited items; it is not OAI-PMH compliant and does not use open source software although it also has an intention to become an OA repository. It archives master thesis and PhD thesis, but have the intention to archive other material types in future (journal articles, books etc.). The only Croatian repository that meets full definition of OA repository, and the only one that archive journal articles, is the *University of Zagreb Medical School Repository* (837 items). Full texts of all of the items are freely available and the repository is OAI PMH compliant. Open source software EPrints, version 2.3.13.1 was chosen and adapted to meet the institution's and the users' needs. Types of archived material are journal articles (658 items), thesis (170), book sections (6), conference or workshop items (2) and one book.

Comparing Croatian practice to those in some other European countries, we have to have in mind, beside other characteristics, the size of the countries. Table 1 shows number of OA repositories in the seven countries (according to the DOAR database), number of OA IRs in the same countries and the population of the countries.

Table 1: Number of OA repositories and OA institutional repositories in seven European countries compared to the population of the countries

	OA repositories	OA IRs	Population
Italy	59	51	61 000 000
Poland	45	37	38 000 000
Greece	14	12	11 000 000
Hungary	11	10	10 000 000
Bulgaria	3	3	7 200 000
Croatia	5	4	4 500 000
Slovenia	3	3	2 500 000

It is obvious that the countries with more inhabitants have more repositories – Italy, for example, with its population of about 61 million, has 59 repositories, Poland has 45 repositories, and the smallest countries, Croatia and Slovenia, have 5 and 3 Open Access repositories. It is also obvious that the majority of the repositories in all the countries are institutional repositories and in the countries with the smallest number of repositories – Bulgaria and Slovenia – all of the repositories are institutional. That shows the importance of institutions (mainly scientific institutions) in setting up Open Access repositories. The institutions and their libraries are aware of their users' needs, they know that the repositories can raise impact of research results of the institution.

Table 2 shows five most usual material types in OA repositories in the seven European countries. All the countries have repositories with journal articles

(that is the most common material type), books or chapters and conference papers. Almost all the countries have repositories with theses and learning objects.

Table 2: Material types in repositories in seven European countries

	Journal articles	Books or chapters	Conference papers	Theses	Learning objects	Total no of repositories
Italy	37	24	34	39	9	59
Poland	36	34	7	14	13	45
Greece	8	4	4	7	2	14
Hungary	2	3	3	2	1	11
Croatia	3	2	2	3	-	5
Slovenia	2	1	2	2	1	3
Bulgaria	3	1	1	-	1	3

When analysing software used for setting up and operating OA repositories, we have to have in mind that numerous Open Access initiatives recommend usage of open source software (such as EPrints and dSpace). Another recommendation is OAI-PMH compliance (Open Access Initiative – Protocol for Metadata Harvesting) that is important for ensuring better visibility and interoperability of OA repositories.

Table 3: Software used in OA repositories in the seven European countries

	EPrints	dSpace	Other	Total no of repositories
Italy	28	21	10	59
Poland	2	3	40	45
Greece	-	9	5	14
Hungary	8	2	1	11
Croatia	2	1	2	5
Slovenia	1	-	2	3
Bulgaria	1	2	0	3

Majority of repositories in the seven countries use EPrints or dSpace software (Table 3). The only country with significant use of other software is Poland where 34 repositories use dLibra software that emerged from a Polish national project for building digital libraries (the software is OAI-PMH compliant).[20]
Another important issue when comparing OA repositories in the seven countries is language of deposited material. Official languages in the countries are not understandable to a great number of people (although some languages are "larger" than others). That is the reason why scientists in the countries sometimes publish their research results in English. As can be seen from the Table 4, in four of seven countries there are more repositories with the items written in

---

[20] DLibra : digital library framework. http://dlibra.psnc.pl/index.php?lang=en (accessed May 12, 2011).

the countries' official languages – 49 of Italian 59 repositories have texts in Italian; 43 of Polish 45 repositories have texts in Polish; all the 14 Greek and all the 5 Croatian repositories have texts written in their countries' official languages (i. e. in their scientists' mother tongue). The second language represented in the repositories is English. All the countries have at least two repositories (Slovenia) that deposit material written in English. Other languages that are represented in the repositories are French (in 3 Italian and one Polish repository), German (in two Polish repositories), Spanish (in two Italian repositories), Czech (in one Hungarian repository), Russian (in one Hungarian repository), Armenian (in one Polish repository) and Arabic (in one Italian repository).

Table 4: No. of repositories with items in different languages

	Official language of the country	English	Other	Total no of repositories
Italy	49	40	6	59
Poland	43	16	4	45
Greece	14	10	2	14
Hungary	2	8	1	11
Croatia	5	3	0	5
Slovenia	2	2	0	3
Bulgaria	1	3	0	3

Conclusion

Literature review proves that OA IRs can increase impact and visibility of a country's research output. It is very important when a country is scientifically peripheral – research results published by its scientists are not always visible and accessible to the global scientific community. If we analyse some of European repositories, we can see that the majority of them are institutional repositories, set up mainly at universities or some other scientific institutions. They are usually set up on librarians' initiative. It proves the importance of scientific institutions in preserving and making available their research output. It also proves the importance of libraries that are the first place where access problem has to be solved if the users needs want to be met.

The majority of the OA repositories in analysed countries preserve journal articles, conference papers, theses and books. They usually use open source or other software that ensures interoperability and increase visibility of the OA repositories. Majority of repositories in four countries deposit material written or produced in their official languages but there is also a significant number of repositories that deposit materials in English.

Some of the analysed countries do not use all the possibilities OA offers. They should set up more institutional repositories as their research results could be more visible and accessible outside the countries. The worst situation is in Bulgaria, but the number of OA repositories in Croatia and Slovenia is also far too small (although the countries are smaller than other analysed countries). Institu-

tions and their libraries should try to organize working groups to make project proposals for setting up institutional repositories. Those working groups should have in mind the size and the organizational structure of the institution as well as scientific fields covered by the researchers of the institution. The groups should propose the most appropriate software and material types that could be deposited in a future repository. Members of working groups should be librarians, scientists, management, policy makers, technical staff etc. Setting up an OA repository is not easy; it includes short term and long term planning. But the results could be positive – research impact of the institution, its scientists and students can be improved.

References

Association of Research Libraries. The case for institutional repositories: a SPARC position paper. 2002. http://www.arl.org/sparc/bm~doc/ir_final_release_102.pdf (accessed May 02, 2011).

Besemer, Hugo. Gathering evidence about the effectiveness of "Open Access" publishing policies in agriculture. 2006. http://agriscontent.wordpress.com (accessed May 05, 2011).

Budapest Open Access Initiative. http://www.soros.org/openaccess/read.shtml (accessed May 02, 2011).

DLibra: digital library framework. http://dlibra.psnc.pl/index.php?lang=en (accessed May 12, 2011).

DOAR: Directory of Open Access Repositories. www.opendoar.org (accessed May 11, 2011).

Eysenbach, Gunther. Citation advantage of Open Access articles. PloS Biology. 2006. http://biology.plosjournals.org/archive/1545-7885/4/5/pdf/10.1371_journal.pbio.0040157-S.pdf, (accessed May, 10, 2011).

FOI dlib. http://dlib.foi.hr/handle/10439/1 (accessed May 11, 2011).

Harnad, Stevan et al. The access/impact problem and the green and gold roads to Open Access. // Serials Review 30, 4(2004), available at: http://eprints.ecs.soton.ac.uk/10209/1/impact.html (accessed May 02, 2011).

Harnad, Stevan; Brody, Tim. Comparing the impact of Open Access (OA) vs. non-OA articles in the same journals. // D-lib Magazine 10, 6(2004), available at: http://www.dlib.org/dlib/june04/harnad/06harnad.html (accessed May 02, 2011).

Harnad, Stevan; Hajjem, Chawki. The Open Access citation advantage: quality advantage or quality bias? 2007. http://eprints.ecs.soton.ac.uk/13328/1/moed.html (accessed May 05, 2011).

Hebrang Grgić, Ivana. Open Access to scientific snformation in Croatia: increasing research impact of a scientifically peripheral country. Saarbrücken: Lambert Academic Publishing, 2011, p 54.

International Monetary Fund. Transition economies: an IMF perspective on progress and prospects. http://www.imf.org/external/np/exr/ib/2000/110300.htm#I (accessed May 05, 2011).

Lynch, Clifford A. Institutional repositories: essential infrastructure for scholarship in the digital age. // Libraries and the Academy, 2, 3(2003), pp. 327-336.

Marušić, Ana; Marušić, Matko. Small scientific journals from small countries: braking from a vicious circle of inadequacy. // Croatian Medical journal 40, 4 (1999), pp. 508-514.

Moed, Henk F. Statistical relationships between downloads and citations at the level of individual documents within a single journal. Journal of the American Society for Information Science and Technology 56, 10(2005), pp. 1088-1097.

Open Access in Southern European Countries. http://www.accesoabierto.net/sites/accesoabierto.net/files/OASouthEurope.pdf (accessed May 11, 2011).

ROAR: Registry of Open Access Repositories. 2010. http://roar.eprints.org (accessed May 11, 2011).

Schwartz, Greg J.; Kennicutt, Robert C. Demographic and citation trends in Astrophysical Journal papers and preprints. 2004. http://www.publingresearch.net/Citations-SummaryPaper3_000.pdf (accessed May 05, 2011).

Shin, Eun-Ja. Do impact factors change with a change of medium: a comparison of impact factors when publication is by paper and through parallel publishing. // Journal of information science 29, 6(2003), pp. 527-533.

Stojanovski, Jadranka; Petrak, Jelka; Macan, Bojan. The Croatian national Open Access journal platform. // Learned Publishing 22, 4(2009), pp. 263-273.

The Role of National Citation Index in the Evaluation of National Science

Aleksandra Popovic
University in Belgrade, University Library "Svetozar Markovic"
Bulevar kralja Aleksandra 71, Belgrade, Serbia
popovic@unilib.bg.ac.rs

Sanja Antonic
University in Belgrade, University Library "Svetozar Markovic"
antonic@unilib.bg.ac.rs

Dragana Stolic
University in Belgrade, University Library "Svetozar Markovic"
stolic@unilib.bg.ac.rs

Summary

This paper presents the national citation index as a relevant source for development and evaluation of national science, whose use and impact mark constant increase. Serbian citation index (SCIndeks), open access service for Serbian scientific journals, established in 2005, started in cooperation of the Centre for evaluation in Education and Science (CEON) and National Library of Serbia, and financed by Ministry of Education and Science. It contains more than 1.4 million of references from more than 120.000 articles (more than 40.000 in full text). Journals from natural, biomedical and technical sciences are referred from 2000, social sciences and humanities from 1991. From 2008, by law regulations, citations from Serbian citation index are taken as valid for evaluation of scientific work, beside Web of Science.
Since the social sciences and humanities from Serbia are rarely cited in Web of Science (much less than natural, biomedical and technical sciences), Serbian citation index is covering that part of scientific work. It contains also the bibliometric reports for referred journals with detail and valuable data about characteristics of journals (type of articles, age of references etc). Serbian citation index surely contributed to the general visibility of Serbian science as a part of process in which the authors evaluate journals publish in them and raise again their quality.
Serbian citation index has 7000-9000 visits per day and more than 13.000 registered users. Monthly, it is access app. 200.000 times (comparing with Web of Science - 22.000 accesses per month).

University Library in Belgrade takes a significant role in dissemination of information about Serbian citation index by organizing workshops and courses for academic environment.

Key words: citation index, open access, scientific evaluation, Serbia

Introduction

Small scientific communities, such as Serbian, give additional effort to make results of its scientific work visible to the world science. Some scientific areas (like biomedical) are more perceptible outside the country, but some other disciplines, directed to the local subject areas (history, geology, etc), are rarely presented in the world. The question that can be posed is: how these scientific areas can be evaluated properly? Are they going to be deprived for the possibility of citation analysis which enables the more precise results and which can be easily checked by the rest of the scientific environment? How they can measure and see the actual *impact* of their work to other scientists?

Serbian scientists are present in the citation index – Web of Science, not only through the papers published in the foreign scientific journals. In this moment, 20 journals from Serbia are referred in Web of Science, and 18 among them have Impact Factor. Most of them are from biomedical and technical sciences, and only two from social sciences (psychology and economy). It leaves the great number of national scientific journal not visible in this database. How this gap can be over passed?

The role of Serbian citation index

Serbian citation index is an open access service established in 2005. It is a national service whose goal is to index Serbian scientific journals – journals categorized as periodical scientific publications.

It is developed in cooperation of Centre for Evaluation in Education and Science and National Library of Serbia, and it is financed by Ministry of Education and Science of the Republic of Serbia (Centre for Evaluation in Education and Science, access 2011).

It covers 357 journal titles including cited references. From that number 131 are indexed in full text. Serbian scientists and others have access to more than 120.000 articles, from which more than one third are available in full text. Finally, more than 1.4 million of cited references in these articles are searchable for the citation analysis of Serbian scientists.

Journals from natural, biomedical and technical sciences are referred from year 2000, and social sciences and humanities from 1991 (Serbian citation index, access 2011).

From 2008, bylaw regulation, Serbian citation index is taken as relevant source for evaluation of scientific work, beside Web of Science. In official Regulation of the Ministry of Education and Science (March 2008) is quoted: "The citation

is shown by the total number of citations, hetero citation and fractional, considering the contribution of each author in cited work. Citation from Web of Science and national citation index should be presented separately (Rules of Procedure and quantitative way…, access 2011)."

From the point of view of users, Serbian citation index can fulfill several needs:

- new and accurate information of trends in particular science (subject search)
- receiving the full text for one third of the references (open access)
- information about citations in national journals (citation search)

The open access policy is particularly significant, not only for the subject search (making scientific results easy to get), but also for evaluation of particular journal. From 20 journal referred in Web of Science, 18 are indexed in Serbian citation index and available in full text (KoBSON, access 2011). That offers the possibility for the rest of scientific community to has quick access to the results and trends that are visible in the world science, to use them and cited them properly, and finally to publish in those journals, raising their quality again.

Serbian scientists in Web of Science and Serbian citation index

Comparing the results in both indexes for the particular authors (social sciences and humanities), it can be noticed the significant difference between the number of citations in Web of Science and Serbian citation index (Elleby Anita, 2010). Vladeta Jerotic is psychiatrist and writer, Ranko Bugarski is linguist and Dusan Ivanic is philologist.

Table 1: Citation in Web of Science and Serbian citation index

Author	Citations in Web of Science (from 1996)	Citations in Serbian citation index (from 1991)
Jerotic, Vladeta	3	143
Bugarski, Ranko	36	252
Ivanic Dusan	2	56

This unbalance between the results from two citation indexes shows that very productive and important authors, especially from social sciences and humanities, can have small number of citations in Web of Science, but much more in national citation index. That takes us to the observation that lack of citations in Web of Science doesn't mean that the work of specific author is without impact, but that it should be observed within the different context.

Information about journals

Serbian citation index offers the important bibliometric information about the particular journal which is indexed, beside detailed bibliographic data (with classification). Serbian citation index has a Journal Bibliometric Report which

contains information like: number of citations, number of hetero citations, share of foreign authors, number of references, age of references, etc.
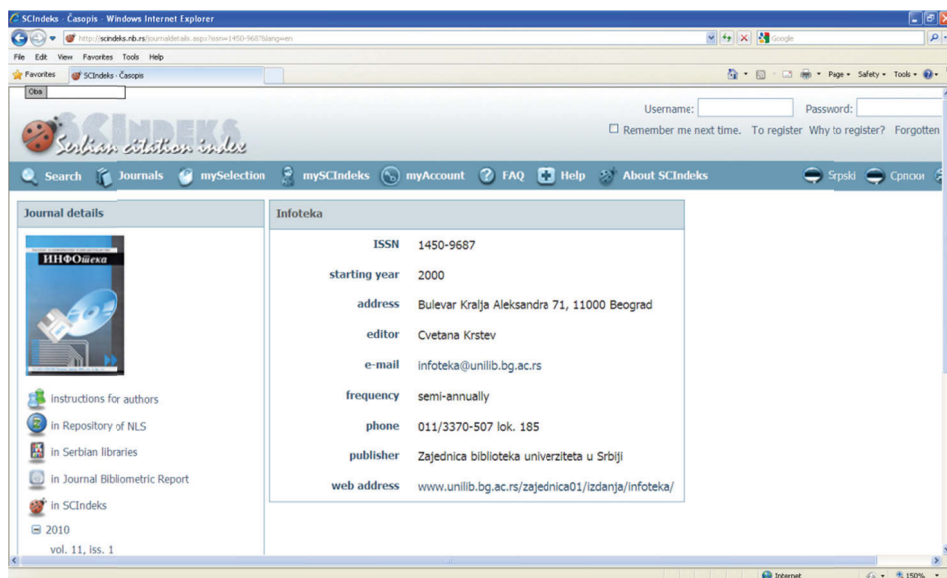
Picture 1: Web page of Serbian citation index - Information about journal and links to other resources: shared catalogue, repository and Journal Bibliometric Report (http://scindeks.nb.rs/journaldetails.aspx?issn=1450-9687)
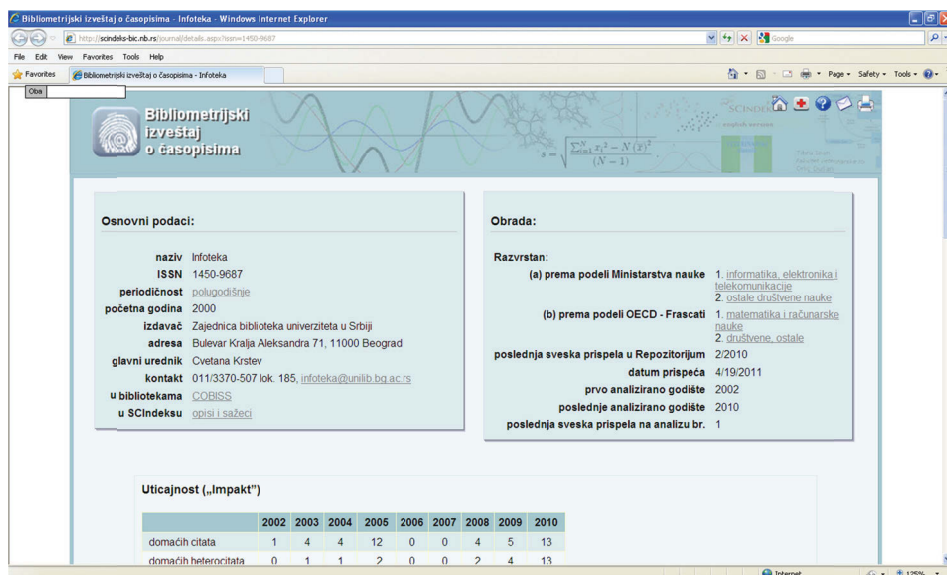
Picture 2: Part of Infotheca Bibliometric Report:bibliographic data and classification, (http://scindeks-bic.nb.rs/journal/details.aspx?issn=1450-9687)

Uticajnost („Impakt")

	2002	2003	2004	2005	2006	2007	2008	2009	2010
domaćih citata	1	4	4	12	0	0	4	5	13
domaćih heterocitata	0	1	1	2	0	0	2	4	13

- domaćih citata
- domaćih heterocitata

Napomena: Međunarodna citiranost, pa tako ni impakt faktori časopisa se ne prikazuju, pošto rezultati nisu uporedivi po godištima. U prethodnim izveštajima ona je, iz tehničkih razloga, utvrđivana na osnovu CD izdanja ISI citatnih indeksa i prikazivana kumulativno (tzv. trogodišnji citatni prozor), dok se od

Picture 3: Impact of Infotheca (domestic citation and domestic hetero citation)

Bibliometrijski kvalitet

	2002	2003	2004	2005	2006	2007	2008	2009	2010
udeo (%) legitimnih radova	100.00	100.00	41.00	100.00	100.00	100.00	100.00	100.00	100.00
opremljenost radova	9.22	10.00	10.00	10.00	10.00	6.00	3.83	8.44	9.00
udeo (%) radova inostranih autora	0.00	47.37	47.06	0.00	18.18	12.50	22.22	44.44	50.00
udeo (%) radova na svetskim jezicima	0.00	36.84	64.71	53.33	36.36	0.00	0.00	100.00	100.00
prosečan broj referenci u radovima	8.44	3.57	8.81	6.60	4.70	7.67	28.00	15.75	18.75
udeo (%) radova s bar jednom referencom	88.88	52.63	81.81	73.91	90.91	37.50	66.67	88.89	100.00
udeo (%) časopisnih referenci	36.84	36.76	23.71	54.55	21.28	65.22	17.86	19.05	17.33
udeo (%) referenci na stranim jezicima	89.47	64.70	88.65	62.12	51.06	91.30	64.29	88.10	78.67
prosečna starost referenci	3.18	5.45	3.33	4.85	4.85	2.00	6.33	7.38	8.38

- legitimnih radova
- opremljenost radova
- radova inostranih autora
- radova na svetskim jezicima
- prosečan broj referenci
- radova s bar 1 referencom
- časopisnih referenci
- referenci na stranim jezicima

Picture 4: Table and graph about Infotheca bibliometric quality

Information about journals can be very valuable to the authors who can see the impact and scientific features of the journal, can evaluate it and finally decide if they should publish in it. This is the process that brings to the new raise of journal quality (Li Jiang, 2010).

Use of Serbian citation index and education

The use of Serbian citation index marks constant increase. It has 7000-9000 visits per day, and 200.000 access per month (2 million per year). Registered users (more than 13.000) have possibility to use service "My account" for additional preferences.

Still, for the much of academic population Serbian citation index is unknown. That is why the University librarians include Serbian citation index in its educational program for each group that participate the organized courses. The Department for scientific information and education, as its regular activity, has dissemination of information about all electronic resources available to members of University of Belgrade, and others. Some of those activities are: practical courses for the students of library and information sciences, courses for PhD students about electronic resources acquired through KoBSON (Serbian Library Consortium for Coordinated Acquisition), presentation on searching electronic resources for professors and students at different faculties (Filipi Matutinovic Stela, 2011), once a week user education about library skills and information skills. The presentation of Serbian citation index is obligatory part of all these educational programs (University Library, access 2011).

Conclusion

Web of Science covers 10-12% of scientific world journals. The great part of world science stays outside of it. It should not be expected that the rest of journals (especially national journals oriented toward specific characteristics of particular country) will be visible to the world science through this citation index.

The solution for the problem (how to cover by citation analysis all scientific areas, with international and national orientation, in natural, biomedical and technical sciences as well as social sciences or humanities), can lie in the development of national citation indexes. The role of those indexes is not competitive, but rather additional. The example of Serbian citation index shows that the way is not to subordinate national science to international, but to analyze its particular qualities and to measure it within its proper environment.

References

Centre for Evaluation in Education and Science (CEON). http://ceon.rs (access date 19.06.2011)

Elleby, Anita; Ingwersen, Peter. Publication point indicators: A comparative case study of two publication point systems and citation impact in an interdisciplinary context. // *Journal of Informetrics*. volume 4 (2010), number 4, pages 512-523

Filipi Matutinovic Stela. Elektronski izvori informacija u nauci: značaj, vrste, dostupnost, procena vrednosti = Scientific electronic information sources: significance, availability, valuation. 23. May 2011. http://www.unilib.bg.ac.rs/edukacija/index.php (access date 21.06.2011)

KoBSON = Serbian Library Consortium for Coordinated Acquisition
http://kobson.nb.rs/nauka_u_srbiji/referisani_casopisi.23.html (access date 23.08.2011)

Li, Jiang; Sanderson, Mark; Willett, Peter; Norris, Michael; Oppenheim, Charles. Ranking of library and information science researchers: Comparison of data sources for correlating citation data, and expert judgments. // *Journal of Informetrics.* volume 4 (2010), number 4, pages 554-563

Pravilnik o postupku i načinu vrednovanja, i kvantitativnom iskazivanju naučnoistraživačkih rezultata istraživača = Rules of Procedure and quantitative way of evaluating and reporting on the results of scientific research (access date 24.08.2011)

Serbian Citation Index-SCIndeks. http://scindeks.nb.rs (access date 19.08.2011)

University Library "Svetozar Markovic". http://www.unilib.bg.ac.rs (access date 20.06.2011)

Webometric Ranking of University of Zagreb and Other World Universities

Krešimir Pavlina
Department of Information Sciences,
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
kpavlina@ffzg.hr

Summary

This paper shows the shortcomings in the implementation of the algorithm used by Aguillo et al. for ranking of world universities, which negatively influenced ranking of 768 world universities, including the University of Zagreb. By correct application of the algorithm, ranking of University of Zagreb extremely improves, so the University of Zagreb is ranked 174th in the world (as opposed to 1355th by Aguillo et al.), 41st in Europe, 3rd in Central and Eastern Europe and 1st in Croatia.

Key words: webometrics, university ranking, University of Zagreb

1. Introduction

There are several algorithms used for ranking world universities. Rankings are usually calculated using combination of factors such as the number of Nobel prize winners who were educated or had worked at the university at the time of obtaining award, number of scientific papers published in selected journals, number of employed highly cited scientists, interest of potential students for enrollment or results from quality control survey.[1][2]

Usually universities with better funding have advantage in these rankings because they have greater financial resources which can attract highly cited scientists and candidates for the Nobel Prizes.[3]

In year 2004 Aguillo and colleagues launched the Webometrics Ranking of World's Universities (www.webometrics.info) [4], the ranking of universities by their presence on the Internet with the aim of promoting and enriching the content of university web pages.

2. Webometrics ranking algorithm

Aguillo et al. collected results from web search engines (Yahoo, Google Search, Google Scholar) for domain of each university and ranked universities according to collected results.

Algorithm for ranking world universities, used by Aguillo et al., is based on four basic parameters.

2.1. Visibility

Number of web pages linking to university web site, indirectly shows the relevance of university web pages. Universities are ranked from the most linked, the less linked.

2.2. Number of university web pages

Number of web pages that are crawled by web search engines (Google, Yahoo) on university domain shows the size of the university web site.

2.3. Rich files

Number Adobe Acrobat (.pdf), Adobe Postscript (.ps), Microsoft Word (.doc) and Microsoft PowerPoint (.ppt) documents indexed by Google's search engine at the university domain. Number of documents represents the amount of educational material that the university published.

2.4. Scientific papers indexed in Google Scholar

Number of scientific papers containing the name of the university, which is indexed by Google Scholar.

2.5. Total rank

When calculating total rank different weight is given to this parameters so visibility bring 50%, size of Web pages 20%, rich files 15% and scientific papers 15%.

Diagram 1. Graphical visualization of weight factors given to each parameter
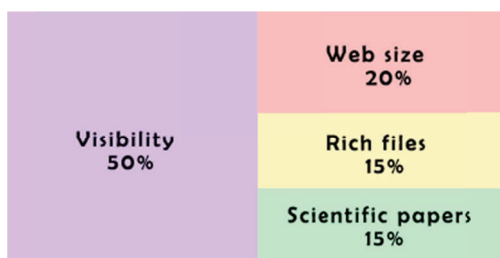
3. Methodological inconsistencies conducted by Aguillo et al. in ranking of world universities

Aguillo et al. insist on using the results from only one domain for every university, and claim that the use of multiple domains is a bad practice that should be eradicated. They claim that it is easier to achieve greater recognition of the university and that it will allow better navigation for foreign students while searching for information at university web pages.

They also claim that the use of multiple domains reduces the visibility factor of the university on the World Wide Web. Using multiple domains does not diminish the visibility factor, because it is possible to sum results of all domains of certain university.

Information systems should support and facilitate real life, the introduction of a centralized domain would impede operations at the University of Zagreb. Each faculty of the University of Zagreb now has its own domain, e.g. domain Faculty of Humanities and Social Sciences is ffzg.hr, while domain of the Faculty of Electrical Engineering and Computing is fer.hr. By pooling all the domains of faculties under central university domain, faculties would be represented by subdomains under unizg.hr domain, so the previous two faculties would be available on the subdomains ffzg.unizg.hr and fer.unizg.hr. It would probably lead to the use of these subdomains in e-mail addresses, so, for example, my email address would no longer be kpavlina@ffzg.hr but kpavlina@ffzg.unizg.hr. Hereby, we would virtually eliminate the possibility of using faculty subdomains because it would certainly be more difficult to remember the address forum.ffzg.unizg.hr instead of the forum.ffzg.hr.

The introduction of a central university domain would aggravate life for all students, teachers and other regular visitors to faculty web sites, because they would inevitably have to remember one more domain in address of web pages of their faculties, under the pretext that we want to enhance navigation to visiting foreign students who are only a minority users which access university web pages.

It is interesting to note that Aguillo et al., explicitly state that they do not want to sum results of all domains of faculties of the University of Zagreb, while for "John Hopkings" university, University of Manchester, Technical University of Munich, Harvard University, Cardiff University, University of Barcelona, Kansas State University, University of Illinois and "Pierre and Marie Curie" university they summed results for several of their domains, according to Aguillo.[5] This proves biased collection of results, because the rule on one domain for every university was applied for some universities, and wasn't applied for other. It is interesting to note that of 12.000 universities ranked by Aguillo et al., 6.4% universities (768), according to their data, used more than one domain.

It is worth to mention that on page www.webometrics.info, Aguillo et al. published the results of their webometric ranking of world universities in a

nontransparent way, because they publish only ranks of the university, rather than absolute results that would allow the verifiability of the results. Ranks are published under the pretext that the results of the search engines are changing daily and that publishing of the absolute results wouldn't allow any comparisons, but there remains no doubt that the publication of only ranks reduced verifiability of results of their research.

4. Methodology
The aim of this research is to determine the ranking of the University of Zagreb and other world universities using webometric research methods used by Aguillo et al.[6] The study was conducted in June 2011 and corrected methodological inconsistencies conducted by Aguillo et al. during the research by not including results for all domains of the University of Zagreb. This research included results from central domain and 29 faculties, 3 academies and 1 subsidiary of University of Zagreb.

Table 1. Domains of constituent units of University of Zagreb

www.agr.hr	www.ffzg.hr	www.pmf.hr
www.arhitekt.hr	www.geof.hr	www.rgn.hr
www.erf.hr	www.gfv.hr	www.sfzg.hr
www.efzg.hr	www.grad.hr	www.sumfak.hr
www.fer.hr	www.grf.hr	www.ttf.hr
www.fkit.hr	www.kbf.hr	www.ufzg.hr
www.foi.hr	www.kif.hr	www.vef.hr
www.fpzg.hr	www.mef.hr	www.adu.hr
www.fpz.hr	www.simet.hr	www.alu.hr
www.fsb.hr	www.pravo.hr	www.muza.hr
www.pharma.hr	www.pbf.hr	www.hrstud.hr

During data collection following number of results was collected for 12.000 universities:
- 961.495.021 web pages
- 816.858.295 links to university web pages
- 67.707.058 rich files
- 192.029 638 research papers indexed in Google Scholar search engine

5. Results

The results show extreme improvement in ranking of University of Zagreb, which is ranked at 174th place in the world, as opposed to the 1355th place where it was ranked by Aguillo et al.[7]

University of Zagreb also achieved a higher rank among European universities, so it is ranked in 41st place of the 5044 European universities, which placed it in the top 1% of European universities.

The success of the University of Zagreb is more significant if we take into account the average expenditure for higher education in countries where universities are located. According to UNESCO data [8], the average investment in higher education for the top 50 universities in Europe amounts to $ 464 per capita, while allocations for higher education in the Republic of Croatia amounted to $ 191 per capita, which is only 41% average. The Republic of Croatia is a country with the lowest financing of higher education between European countries that have a university ranked in the top 50 universities in Europe. Portugal is a country with the next lowest higher education funding ($ 299 per capita), which has a university located in the 50 highest-ranked universities in Europe. Portugal invests 56% more funding on higher education per capita with respect to the Republic of Croatia.

Research results will be displayed by individual variables Aguillo et al. use when ranking universities.

5.1. Number of university web pages

For the University of Zagreb Aguillo et al. have taken only the number of pages available on the university domain (unizg.hr). It is apparent that the stated number of web pages represents only 18.2% of the total number of web pages found on domains of all faculties of the University of Zagreb. Aguillo et al. placed University of Zagreb, according to the number of web pages, at 2583rd place in the world [7], and taking into account the number of web pages found on domains of all faculties of the University of Zagreb, it is placed on 77th place in the world.

Chart 1. Comparison of web size of University of Zagreb
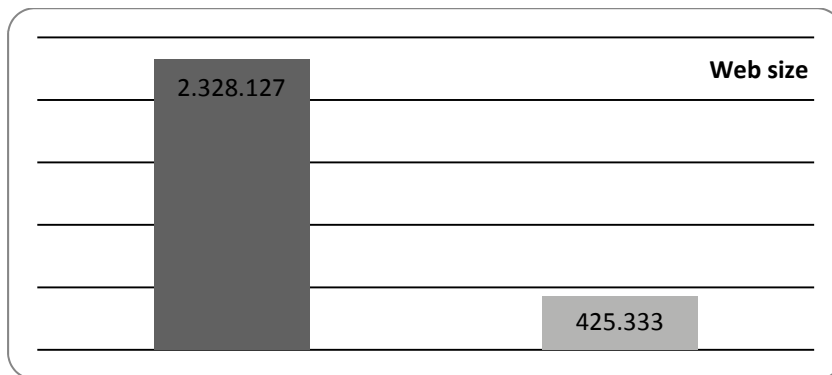
Five domains of the University of Zagreb with the largest number of web pages contribute to the total number of web pages with 82.58%.

5.2. Visibility

To domain unizg.hr refers 81.666 links from other websites, and the 562.913 links point to web pages on domains of all faculties of University of Zagreb. It is evident that Aguillo et al. in the research reduced number of links pointing to domains of University of Zagreb by 85.5%. Visibility have very large infulence in ranking (50% of the total grade), so the rank of University of Zagreb was greatly lower. Aguillo et al. placed, by number of links, University of Zagreb on 1448th place in the world [7], according to this study it is on the 257th place.

Chart 2. Comparison of number of links to University of Zagreb web site(s)
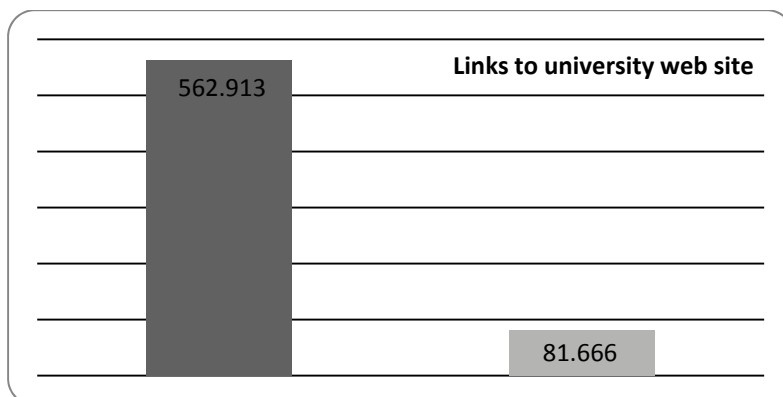
Ten faculties with the highest number of links pointing to their websites, contribute to the total number of links of University of Zagreb with 80.9%.

5.3. Rich files

The total number of rich files found at all domains of the University of Zagreb is 117,714, while only 29.933 files are found at central university domain (unizg.hr). Aguillo et al. in research used only files found on the central university domain (unizg.hr), and they represent only 25.4% of the total number of documents found on the website of all faculties of University of Zagreb. According to the number of rich files Aguillo et al. rank University of Zagreb at 1324th place in the world [7], according to this study University of Zagreb is ranked at 80th place.

Chart 3. Comparison of number of indexed rich files on University of Zagreb domain(s)
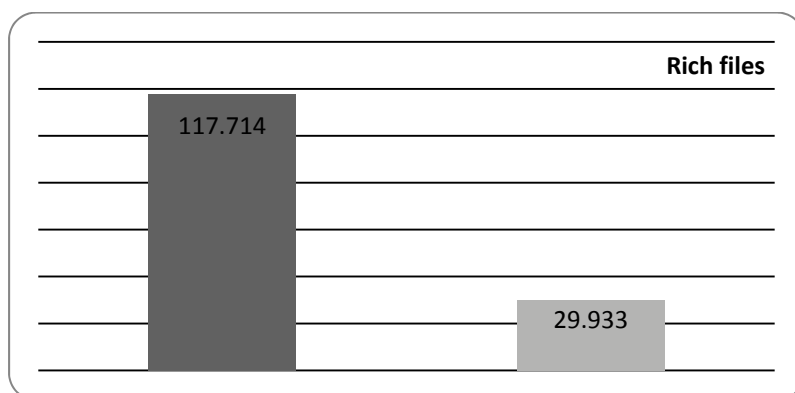
Ten faculties with the highest number of rich files make up for 82.6% of the total number of files of University of Zagreb.

5.4. Scientific paper indexed in Google Scholar

University of Zagreb is represented with 38.767 papers in Google Scholar database. In this parameter there is a large discrepancy with respect to research conducted by Aguillo et al. Aguillo et al. ranked University of Zagreb, according to the number of scientific papers on 1295th place in the world [7], while according to this research it is at 612th place.

5.5. Comparison with world universities

There is a strong correlation in the ranks of leading universities, but it can also be noted that the ranking of University of Zagreb was extremely affected with methodological inconsistencies in research conducted by Aguillo et al., because

it was rated at the 1355th place [7], while by proper application of algorithm it is ranked 174th in the world.

Table 2. Excerpt from the ranking of world universities

Rank in this research	University	Rank in research conducted by Aguillo et al.
1.	Massachusetts Institute of Technology	1.
2.	Harvard University	2.
3.	Stanford University	3.
4.	University of California Berkeley	4.
5.	Cornell University	5.
6.	University of Michigan	7.
7.	University of Wisconsin Madison	6.
8.	Pennsylvania State University	11.
9.	University of Washington	9.
10.	University of Minnesota	8.
	...	
174.	**University of Zagreb**	**1355.**
	...	
225.	University of Ljubljana	205.
	...	
304.	University of Budapest	360.
	...	
763.	University of Belgrade	625.
	...	
2015.	University of Sarajevo	1949.

6. Conclusion

Aguillo et al. started ranking of world universities with the noble intention of raising academic public awareness about the importance of publishing information on the university websites, but unfortunately made several methodological errors in conceptualizing the research and data collection that compromised 768 universities, including University of Zagreb. This study showed that the University of Zagreb is ranked 174th in the world, not 1355th like in research of Aguillo et al., which certainly indicates that a large error in the methodology was done.

The final results of this study will be published in September 2011 at the website unirank.org. There will be published all of the results obtained in this

research and methodology will be described in detail so that independent researchers could repeat the study.

Problem of insufficient public information about the methodologies of ranking universities which often misguided uninformed readers was identified during the research. So this research represents the beginning of the project of creating website unirank.org where it will be possible to compare methodologies of different ranking systems of universities, and where they will be able to find results of ranking of universities according to a number of different rating systems.

References

[1] Academic Ranking of World Universities – Ranking Methodology, (http://www.arwu.org/ARWUMethodology2010.jsp) (1.7.2011.)

[2] HEEACT – Performance Ranking of Scientific Papers for World Universities, (http://ranking.heeact.edu.tw/en-us/2010%20by%20field/page/methodology) (24.7.2011.)

[3] Billaut, J., Bouyssou, D., Vincke, P. Should you believe in the Shanghai ranking?, 2009. URL: (http://hal.archives-ouvertes.fr/docs/00/40/39/93/PDF/Shanghai_JCB_DB_PV.pdf) (1.7.2011.)

[4] Aguillo, I. F.; Granadino, B.; Ortega, J. L.; Prieto, J. A.. Scientific research activity and communication measured with cybermetric indicators. Journal of the American Society for the Information Science and Technology, 57(10): 1296 - 1302., 2006.

[5] Aguillo, I. F. Changes scheduled for the July 2009 edition, blog entry dated 5.3.2009. URL: http://internetlab.blogspot.com/ (1.7.2011.)

[6] Webometrics – Ranking of World's Universities – Methodology. 2011. URL: http://www.webometrics.info/methodology.html (3.7.2011.)

[7] Webometrics – Ranking of University of Zagreb. 2011. URL: http://www.webometrics.info/details.asp?univ=unizg.hr (3.7.2011.)

[8] UNESCO Institute for Statistics – Education, 2011. URL: http://www.uis.unesco.org/ (1.7.2011.)

Standing of the Croatian Language at the European Universities – Web Content Analysis

Marija Roglić

Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, 10 000 Zagreb, Croatia
mroglic2@ffzg.hr

Summary

In the socialist Yugoslavia the official language policy insisted on one language with two standard varieties. The unity of the language was emphasized, interpreting differences among each nations language as factors enrichening the „common language" diversity. Collapse of the Federation signified the breakdown of the „common language". As a result today we talk about standard Bosnian, Croatian, Montenegrin and Serbian.
The aim of this paper is to examine to what extent are these languages acknowledged and studied as separate and different languages at the European universities. Paper presents the results of an analysis conducted at websites of eighty-six universities in twenty seven European countries.

Key words: Bosnian/Bosniak, Croatian, Montenegrin, Serbo-Croatian, Serbian, BCS, BCMS

Introduction

There has been much talk on the position and recognition of the Croatian language. It is and has remained a delicate subject to this day. Some linguists as Mate Kapović (2009: 1) argue that the main reason for the preservation of the name Serbo-Croatian lies in four simple reasons: inertness, scientificness, convenience, and partly viciousness of foreign linguists. The fact is that the knowledge once learned is difficult to change. Also, some scientists to this day regard Croatian and Serbian as one language, ignoring the „local conflicts". Others find easier to treat Bosnian, Croatian, Montenegrin and Serbian as one language, particularly from a methodological point of view. In addition, when quoting older sources, which go by the name Serbo-Croatian, the easiest way is to simply leave it so, both by Serbo-Croatian, Croatian and Serbian sources. Lastly, Kapović states that some may insist on that name for purely political or personal reasons.

According to McGuigan (2011: 1) Serbo-Croatian is more accurately an *umbrella* term for three distinct languages (Bosnian, Croatian, and Serbian) tied

together by similarities. Serbo-Croatian contains loan words from many different languages, as a result of the history rich with various allegiances.

From the Serbian side the language contains many words from both Turkish and Greek. From the Croatian side the language contains many words from German and Latin. The language is written in both Latin and Cyrillic alphabet. For the Bosnian language sometimes it is also used the Arabic alphabet.

Serbo-Croatian was standardized as a single language during the era of Yugoslavia, from 1918 to the breakup of the Soviet Union in 1991. During this period Serbo-Croatian was one of the three official languages, alongside Macedonian and Slovenian. Following the breakup of Yugoslavia, the Serbo-Croatian language broke into its constituent parts, with Bosnian, Croatian, and Serbian becoming distinctly recognized languages. Currently in Montenegro there is a push to have Montenegrin recognized as its own language, as well.

The issue of Serbo-Croatian has become highly politicized, and is in many ways a political issue, rather than a linguistic one. As McGuigan (2011: 1) remarks language is viewed by many people as a sign of cultural and political independence. As a result there has been a push since the breakup of Yugoslavia for each distinct social group to have their own dialect recognized as a distinct language.

The arising problem and the main issue which this paper explores is the result it had and has on the study of the languages in question (formerly one) at European universities. It tries to give an objective analysis of the reach and recognition of the Croatian language and its neighbouring languages as sovereign South Slavic languages. Research itself mainly relies on the information and data collected at the university pages of top five or top three[1] national universities in twenty seven European countries.

Hypothesis

Croatian language in Europe, as well as in the world, its prevalence and significance mainly depends on Croatian Diaspora, its exuberance and engagement. The count is only approximate because of incomplete statistical records and naturalization, but (highest) estimates suggest that the Croatian Diaspora numbers between a third and a half of the total number of Croats. Accordingly, Croatian instructorship will be represented in the regions where there is a greater and more influential number of the Diaspora. The largest emigrant groups are in Western Europe, mainly in Germany, where it is estimated that there are around 450,000 people with direct Croatian ancestry. Consequently Serbian or Bosnian

---

[1] Depending on the total number of universities it is taken a sample of top three or top five universities. In the countries with less then twenty universities we have considered only top three. The author has tried to compare only universities with similar academic interests and a considerable number of students.

will be more represented in those countries where there is a greater number of Serbian, and/or Bosnian Diaspora.

As a result of work emigration in the ex-Yugoslavia the greatest number of Bosnian, Croatian and Serbian Diaspora we find in Germany[2].

Method

This paper researches the reach and significance of both Croatian/Bosnian/ Montenegrin/Serbian language(s) and Croatian language within the European cultural club.

The research is focused on 27 European countries and 86 universities and their faculties with a cathedra or an instructorship of Croatian, Croatian- Serbian, Bosnian/Croatian/Serbian[3] or Bosnian/Croatian/Montenegrin/Serbian[4].

To gain impact results and a detailed region-by region review the countries have been divided into five geographical regions.

All the information used in this research has been obtained from the web pages of the universities and their faculties. The content has been analysed according to the type of instructorship described in the syllabus of the course and defined by its name (Croatian, Croatian-Serbian, BCS or BCMS). The academic background of the foreign-language instructor has also been taken into account as it also influences on the type of speech exercise they will have. And the type of spoken language they will learn. If the instructors are from a Croatian academic background it is natural that the students will hear and talk more Croatian then Serbian, even if the courses are called Croatian-Serbian, BCS, or BCMS. The same goes in the cases when there is a Bosnian or Serbian instructor.

The percentage of all the Croatian sponsored instructorship from the total number of instructorship in Europe where Croatian, Croatian-Serbian, BCS or BCMS is studied has been calculated with the help of a report from the Ministry of Science, Education and Sport written by Staša Skenžić[5].

Analysis

The syllabuses of the Slavonic studies per each university have been analysed with an emphasis on the study of the language. The paper has tried to answer two main questions:

- In which percentage is one or are all languages in question taught at the European universities

---

[2] 158 158 Bosnian, 225 309 Croats and 330 608 Serbs and Montenegrians.

[3] Abb. BCS

[4] Abb. BCMS

[5] Hrvatski jezik u okruženju drugosti; Pregled poučavanja hrvatskoga jezika na stranim visokoškolskim ustanovama

- From all the languages in question what is the percentage of study of the Croatian language as a sovereign and unique language

The countries have been divided into five geographical regions according to authors preference in order to provide more systemized data:

- South-eastern Europe
  - Slovenia, Bosnia and Herzegovina, Macedonia, Bulgaria
- Central Europe
  - Poland, Czech Republic, Slovakia, Hungary, Austria
- Eastern Europe
  - Ukraine, Russian Federation
- Western Europe
  - United Kingdom, Ireland, Belgium, Holland, Portugal, France, Switzerland, Germany, Italy
- Northern Europe
  - Denmark, Finland, Norway, Sweden, Estonia, Latvia, Lithuania

Due to lack of information on their web sites of the universities in Eastern Europe data has been collected from the Croatian Ministry of Science, Education and Sports.

Results

Table 1enumerates instructorships per each language found among top 3 (in South-eastern Europe) or top 5 state universities in each country:

Table 1.

Country	Croatian	Serbo-Croatian	BCS	BCMS
BIH	1	-	1	-
Slovenia	-	1	-	-
Macedonia	1	-	-	-
Bulgaria	-	2	-	-
Czech Republic	3	-	-	-
Slovakia	1	-	-	-
Poland	3	-	-	-
Hungary	2	1	-	-
Austria	1	-	2	-
Russian Federation	1	-	-	-
Ukraine	1	-	-	-
Germany	-	-	5	-
Italy	4	1	-	-
Switzerland	-	1	2	-
Holland	-	-	-	-
Belgium	-	-	-	-
Ireland	-	-	-	-
Portugal	-	-	-	-
France	4	1	-	-
United Kingdom	1	1	-	-
Norway	1	-	-	-

Country	Croatian	Serbo-Croatian	BCS	BCMS
Finland	-	-	-	1
Sweden	-	-	-	-
Denmark	-	-	-	-
Estonia	-	-	-	-
Latvia	-	-	-	-
Lithuania	1	-	-	-

Chart 1 describes the percentage of instructorship per language in relation to the total number of Croatian, Serbo-Croatian, BCS and BCMS instructorships at the European universities:

Chart 1.
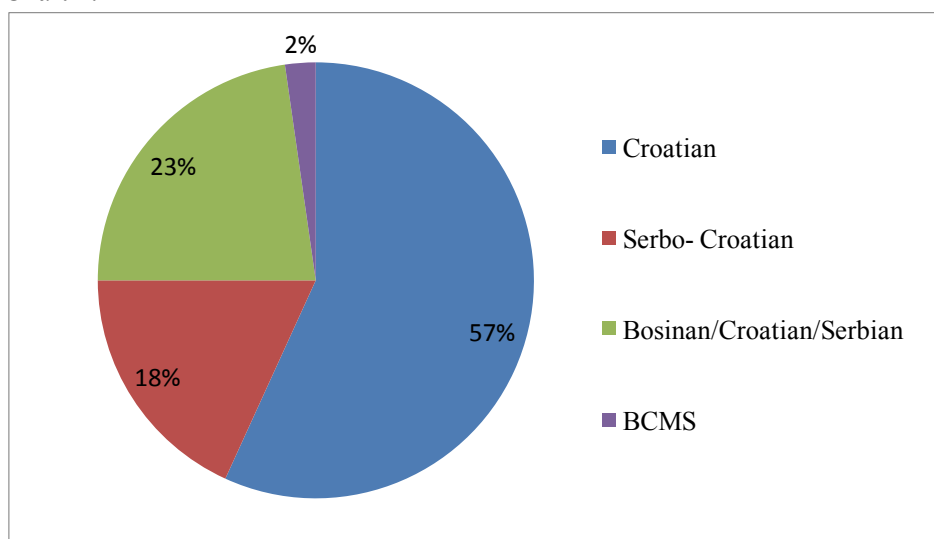
The statistical data obtained with the web content analysis of totally 206 web pages shows that Croatian language is by far the most frequent language used at the foreign instructorship, of all the languages in question.

If we compare these statistics with the report of the Croatian Ministry on foreign instructorship of Croatian language we gain equal results:

Table 2: Crotian instructorship in the World

ARGENTINA
• Universidad de Buenos Aires, Buenos Aires
• Universidad Nacional de Rosario, Rosario
AUSTRIA
• Karl-Franzens University, Graz
BELGIUM
• ISTI - Institute of Translation and translators, Brussels
BULGARIA
• University „Sv. Kliment Ohridski", Sofia
CZECH REPUBLIC
• Univerzity Karlovy, Prague
FRANCE
• Universite de Paris - Sorbonne, Paris
• Universite Jean Moulin, Lyon
• Universite Stendhal, Grenoble
• INALCO Institute, Paris
• Universite Le Miral, Toulouse
INDIA
• University of Delhi, New Delhi
ITALY
• Universita degli Studi di Padova, Padova
• Universita degli Studi di Firenze, Florence
• Universita „La Sapienza", Rome
CHINA
• Beijing Foreign Studies University, Beijing
LITHUANIA
• Vilniaus universitetas, Vilnius
HUNGARY
• Tudomanyegyetem Eotvos Lorand, Budapest
• Berszenyi Daniel, Szombathely
• University „Janus Panonius"Pecs
• Eotvos Jozsef Foiskola, Baja
MACEDONIA
• Sveučiliše „Sv. Kiril and Methodius", Skopje
POLAND
• Uniwersytet Slaski, Sosnowiec
• Uniwersytet Warszava, Warsaw
• Uniwersytet Adama Mickiewicza, Poznan
• Uniwersytet Jagiellonski, Krakow
ROMANIA
• cathedra de Limbi Slave, Bucharest

RUSSIAN FEDERATION
• Faculty of Moscow State University VM Lomonosov, Moscow
U.S.
• Indiana University, Bloomington,
• Indiana University of Iowa, Iowa City
SLOVAKIA
• Comenius University, Bratislava
SLOVENIA
• Faculty of Arts, Ljubljana
SPAIN
• Alcalá de Henares, Madrid
UKRAINE
• Universytet „Taras Shevchenko", Kiev
• Universytet „I. Franko ", Lviv
UNITED KINGDOM
• University of London, London

Conclusion

As Bugarski (2004: 13) concludes we might put forward the view that the linguistic grounds for the disintegration of former Yugoslavia are not to be sought in the mere exsistance of numerous languages at that territory, but rather in the conflicting cultural traditions, national aspirations and political programmes embedded in, and symbolised by, the major languages and their varieties.

Consequently, for the peaceful and stabile region, as well as Europe it is of the crucial importance that all the languages be excepted and acknowledged as soveregin, as it is languages the ones used to mark identity boundaries as well as to express the make-up identity (Byram 2007: 328).

Twenty years from the independence of ex-Yugoslav states it seems we may talk about separate Croatian, Bosnian/Bosniak, Montengrian and Serbian language on an European level without beeing laught at by foreign linguists. As much as it seems unquestionable in the countries of ex-Yugoslavia, it is of the substancial importance to have objective statistical data confirming that these languages, as well as these nations will be acknowledeged as sovaregin and unique in it's identity, language and culture upon entering the European Union. The statistics shown in this paper may assure as in that trend. Represented with 57% we can see that the Croatian language together with the Croatian language policy has sucessfuly struggled to achieve international acceptance. For the Bosinan/Bosinaks and Montenegrians the situation is quite different. Reasons of such vague and versatile acceptance lie in their national language policies which are as versatile, as the statistic of BCS and BCMS language(s) shown in this paper.

References

Bugarski, Ranko. Overview of the linguistic aspects of the disintegration of former Yugoslavia.// *Language in the Former Yugoslav Lands//* Bugarski, Ranko; Hawkesworth Celia. (ed.).Bloomington:Slavica Publishers, 2004, 3-15

Byram, Michael. Exploring Languages, Identity and Education in a European Perspective.// *Zbornik*:*Jezik i identitiet*// Granić, Jagoda. (ed.).Zagreb, Split: Hrvatsko društvo za primijenjenu lingvistiku, 2007, 300-330

Harris, Robert. Evaluating Internet Research Sources. 2010. http://www.virtualsalt.com/evalu8it.htm (2011, march 10)

Huntington, Samuel P. The clash of civilizations and the remaking of world order. New York:

Kačič, Miro. Croatian and Serbian: Delusions and Distortions. Zagreb:Novi most, 1997

Kapović, Mate. Položaj hrvatskoga jezika u svijetu danas.//*Kolo*.01-02 (2009) http://www.matica.hr/kolo/kolo2009_1.nsf/AllWebDocs/Polozaj_hrvatskoga_jezika_u_svijet u_danas (2011, June 12)

McGuigan, Brendan. What is Serbo-Croatian. n.d. http://www.wisegeek.com/what-is-serbo-croatian.htm (2011, June 10)

Statistisches Bundesamt Deutschland. Foreign population on 31.12.2007 by the most frequent citizenships. http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/EN/Content/Statistics/Bevoe lkerung/AuslaendischeBevoelkerung/Tabellen/Content50/TOP10Liste,templateId=renderPrint .psml (2011, June 8)

Skenžić, Staša. Hrvatski jezik u okruženju drugosti. Zagreb: MZOŠ, n.d.

Wikipedia, the free encyclopedia. Croats. http://en.wikipedia.org/wiki/Croats (2011, June 10)

Wikipedia, the free encyclopedia. Differences between standard Bosnian, Croatian and Serbian. http://en.wikipedia.org/wiki/Differences_between_standard_Bosnian,_Croatian_and_Serbian (2011, June 10)

Web References in Graduate Papers at the Department of Information Sciences at the Faculty of Humanities and Social Sciences in Zagreb

Anina Bauer
Department of Information Sciences
Faculty of Humanities and Social Sciences of the University of Zagreb
Ivana Lučića 3, 10 000 Zagreb, Croatia
abauer2@ffzg.hr

Marija Kirin
Department of Information Sciences
Faculty of Humanities and Social Sciences of the University of Zagreb
Ivana Lučića 3, 10 000 Zagreb, Croatia
mkirin@ffzg.hr

Marija Turković
Department of Information Sciences
Faculty of Humanities and Social Sciences of the University of Zagreb
Ivana Lučića 3, 10 000 Zagreb, Croatia
mturkovi@ffzg.hr

Summary

The aim of this article is to give insight into how graduate students at the De-partment of Information Sciences at the Faculty of Humanities and Social Sciences of the University of Zagreb tend to collect their references online. The research is based on the hypothesis, which was drawn up regarding the continuous spread of the Internet use, that the usage of web references in graduate papers has substantially grown in the period from 2004 to 2010. The period in question was chosen according to the availability of digitalised graduate papers in the Digital Repository of the Faculty Library. For this purpose a bibliometric method was used on the sample of over 360 graduate papers in the given period which include graduate papers from Archive Documentation, Library Science, General Information Science and Museology. A detailed analysis of the web references cited in the theses, together with the total number of references, is presented in this article. The analysis includes distinction based on authors and type of online references. One of the aims of the analysis was to investigate which author is most frequently cited even though web references are not always signed by an author. The different types of online references that are used, such as online books, articles, reference materials and others, were taken into

consideration as well. The analysis of the collected data proved the hypothesis that the number of web references increased over the period; however, there is still a great deal of offline based references used in the graduate papers.

Key words: information sciences, web references, graduate papers, bibliometry

Introduction
The research included papers in the period from 2003 to 2010 from the Department of Information Sciences at the Faculty of Humanities and Social Sciences of the University of Zagreb. The period in question was chosen according to the availability of digitalised graduate papers in the Digital Repository of the Faculty Library. The research included graduate papers from the fields of Archive Documentation, Library Science, General Information Science and Museology.
The primary goal was to find out the amount of web references students use in their graduate papers, as well as to see how the trend changed over the years. Besides that, we also wanted to investigate which author is most frequently cited, even though the web references are not always signed by an author, and what type of web references students mostly refer. The work of dr. sc. Đilda Pečarić who researched the Development of Information Sciences in Croatia through a bibliometric analysis of doctoral thesis from the Department of Information Sciences was consulted during the research.

Hypothesis
The research is based on the hypothesis, which was drawn up regarding the continuous spread of the Internet use, that the usage of web references in graduate papers has substantially grown in the period from 2003 to 2010.
Because of the free access to the majority of online content, it is believed that the students refer not only to scientific articles, but also to other types of web references.
In regard to different types of web references, it is predicted that articles from scientific journals are more frequent than monographs since the Faculty Library offers free access to many databases containing scientific journals.

Methods
The research began in November 2010. At the time the list of graduate papers in the library catalogue comprised of 587 papers. It was decided to analyse those papers that are available online in the Digital Repository because of the easier manipulation of digitalised data. The pace of the research was dependant on the availability of the graduate papers in the Digital Repository of the Faculty Library. The papers were taken from the Repository as .pdf files and web references were manually entered into a Microsoft Access database. A bibliometric method was used on those references in order to obtain information regarding

the number of web references. The analysis also includes distinction based on authors and type of web references.

Results

We were able to analyse 362 papers, which contain a total of 6525 references. The average number of references per paper is 18. Paper with the most references contains 70 references, while the paper with the least contains 3 references.

Out of total of 6525 references, 1947 are web references, that is 30 % of the total number (Figure 1.). The average number of web references per paper is 5. Paper with the most web references contains 57 web references, while the paper with the least contains 0 web references. In total, there are 116 graduate papers that did not contain any web references.

Figure 1. The overall percentage of references

The number of papers analysed from each field of study is as following: Archive Documentation (7), Library Science (268), General Information Science (72) and Museology (15) (Table 1). As seen from the aforementioned, the number of papers differs depending on the subject, which has to be taken into consideration while analysing the results for each field separately.

As seen from the Table 1 out of the four fields, the average number of references is the largest in Museology while the smallest is in General Information Sciences. Archive Documentation and Library Sciences have a similar average number of references, 18.86 and 18.84. The arrangement of the average number of web references slightly differs, smallest is in Archive Documentation, in Li-

71

brary Science that number is 5.02 and slightly higher are in General Information Science, 6.78 and Museology, 7.07.

The large delineation of average numbers of web references in Archive Documentation (extremely low) and Museology (high) is the result of a small number of analyzed papers in those fields and, therefore, the averages in question cannot be taken into consideration as representative of the field.

Table 1. The results according to the field per year

field	analysed papers	analysed references	average number of references	analysed web references	average number of web references
Archive Documentation	7	132	18,86	8	1,14
Library Science	268	5048	18,84	1345	5,02
General Information Science	72	993	13,79	488	6,78
Museology	15	352	23,47	106	7,07
total	**362**	**6525**	**-**	**1947**	**-**

The paper from Library Science with the most references has 70 references, from General Information Science 66, and from Archive Documentation and Museology 44 references. Concerning web references, the paper from General Information Science with the most web references has 57, from Library Science 45, from Museology 29 and from Archive Documentation 7 web references.

The average number of web references in General Information Science is considerably high, 6.78. As a possible explanation for the fact that General Information Science has both one of the largest average number of web references and the paper with the highest number of web references is that the papers in this field deal with the subjects more connected to Information Communication Technology than other fields.

Table 2. The number of references in the papers from 2003 to 2010

year	analysed papers	analysed references	average number of references	analysed web references	average number of web references
2003	4	88	22,00	37	9,25
2004	71	1104	15,55	264	3,72
2005	45	794	17,64	175	3,89
2006	69	1089	15,78	276	4,00
2007	63	1137	18,05	346	5,49
2008	78	1605	20,58	525	6,73
2009	27	586	21,70	276	10,22
2010	5	122	24,40	48	9,60

One of the reasons for the unequal number of papers per year is the sporadic input of papers in the Digital Repository of the Faculty Library, meaning the record of the papers was in the Repository but some papers were unavailable for download. Taking that into account, different number of graduate papers (per year) were analysed and the results may vary. Because of the small number of papers available for analysis, the results for the years 2003 and 2010 were not taken into consideration. The most accurate results are for the years 2004 and 2008 because of the large number of papers written in those years (Table 2).

The average number of web references as seen from Table 2 proves the hypothesis that the usage of web references increases over the years.

The graphical representation of this is Figure 2 which shows both the change of all references and the change of web references from 2004 to 2009.
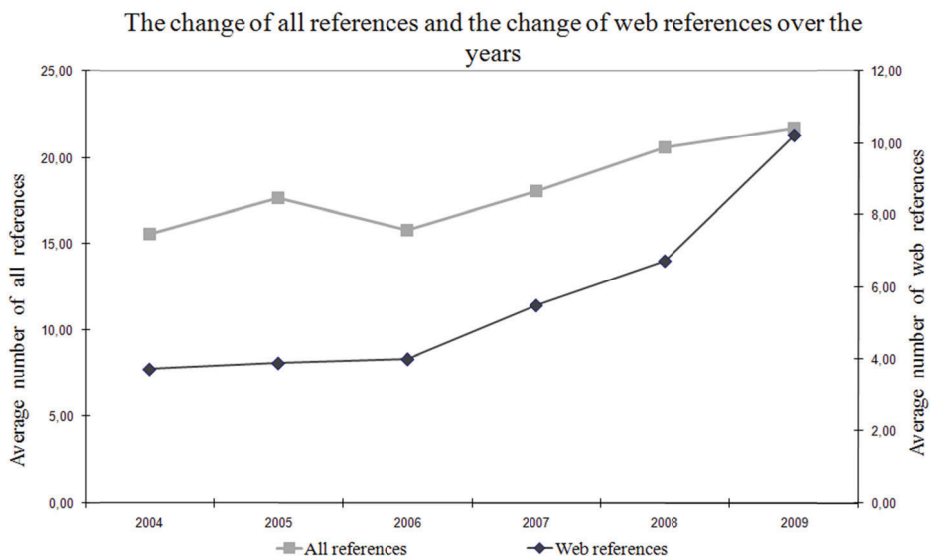
Figure 2. The change of all references and the change of web references over the years

Besides the frequency of the web references, we also wanted to find out what types of web references are most frequently used. The majority of web references used include web pages, reference materials and articles. Students mainly refer to web pages which may contain some information about the organization the paper refers to, such as libraries, universities and other institutions. Secondly, they refer to reference materials - out of which Wikipedia is the most-cited encyclopaedia with 102 references which make 5.24% of all web references; and Narodne novine (The Official Gazette) the most-cited newspapers, from which laws, standards, guidelines etc are referred, with 67 references, which make 3.44% of all web references. Thirdly, the graduate papers cite arti-

cles that appear either on the web pages or digitised and digital journals. Also, some students referred to scientific portals instead of actual articles. There is a small amount of online books used as references. Other web references include different types of manifests, guidelines, published lectures and alike.

The biggest problem we were faced with was the availability of the graduate papers which, as aforementioned, was limited. There were problems when analyzing the web references since a substantial number of the citations themselves are in some way deficient. For instance, some elements, such as titles, are missing or incorrectly cited (for example subtitles formatted as titles). Some of the citations only have the URL address listed and the date of access often omitted. Typographical mistakes are also common, even within the URL address, which indicates that, in some cases, the addresses were not directly copied from the Internet browser. Although the Department of Information Sciences does offer citation guidelines, the frequency of incorrect citations suggests that the students rarely follow them. Perhaps instead of guidelines a standard for citation should be prescribed for graduate papers.
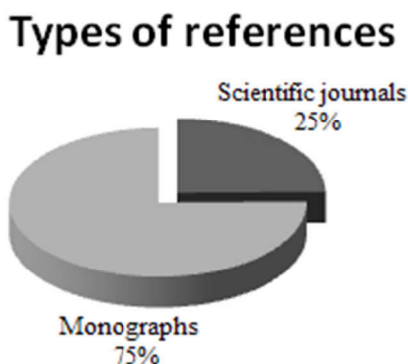
Figure 3. Types of references

Because of the listed problems, as well as because some of the URL addresses do not work anymore, it was not possible to identify the type of reference for 36.67% of all web references. The overall percentage of monographs and scientific journals from the identified types is shown in Figure 3. More referenced are monographs. This is probably so because of the vast number of standalone papers that can be found online such as different types of manifests, guidelines, published lectures and alike.

Concerning authors, 1422 web references, that is 73%, do not have an author listed, while the most commonly cited author in the other 524 web references is doc. dr. sc. Sonja Špiranec with 10 references, followed by prof. dr. sc. Tatjana Aparac-Jelušić with 7 and mr. sc. Jadranka Stojanovski with 6 references.

Conclusion

The analysis of the collected data proved the hypothesis that the number of web references in the graduate papers of students of information sciences at the Faculty of Humanities and Social Sciences increased over the period; however, there is still a significant number of offline-based references used in the graduate papers.

As was presented in the results, the papers with the most references, and particularly web references, were from the field of General Information Science. That is possibly because of the fact that the field deals with the subjects most connected to Information Communication Technology. The analysis also showed that the most-used web references are in fact monographs, even though the scientific journals were expected to be more frequent. Other unforeseen results include the vast usage of Wikipedia articles, which is thought to be uncommon in graduate papers, and the alarming frequency of incorrect citations. It was interesting to find that one of the listed hypotheses was not accurate, however the primary hypothesis that the usage of web references has substantially grown was proven.

A possible prediction based on the results may be that the usage of web references by (graduate) students will continue to grow. It can also be concluded that more accurate results will be available when all of the graduate papers from the given periods are inserted into the Digital Repository of the Faculty Library, which indicates that the conducted research can easily be continued.

References

Pečarić, Đ. Razvoj informacijskih znanosti u Hrvatskoj: bibliometrijska analiza doktorskih disertacija iz informacijskih znanosti 1978.-2007.: doktorska disertacija. Zagreb, 2010

Graduate papers at the Department of Information Sciences at the Faculty of Humanities and Social Sciences of the University of Zagreb ; downloaded from Digital Repository.

Access database "Web reference u diplomskim radovima odsjeka za Informacijske znanosti.mdb" containing Web references in graduate papers at the Department of Information Sciences at the Faculty of Humanities and Social Sciences in Zagreb.

Archival Heritage of Religious Communities in Croatia and Possibilities for its e-Accessibility

Dražen Kušen
State Archives in Osijek
31000 Osijek, K. Firingera 1
drazen.kusen@dao.hr

Summary

Archival heritage of religious communities in Croatia is kept in private archives of the religious communities, Croatian State Archives in Zagreb, other regional state archives', and special archives' of certain public institutions. The number and variety of the religious communities in Croatia, especially in its northern parts, is a reflection of the historic migrations of the population, which shaped a multicultural and multireligious social context. Within that context, during history, the religious communities engaged in a wide range of activities, so the archival materials, created as a result of their work, are a rich source of information in different fields. In accordance with specific functions of the religious institutions (the creators), archival materials containing specific types of information were created. The materials of secular provenance often do not contain such information. Therefore, the type and contents of those pieces of information shape special informational character of the religious archives that keep and provide access to the mentioned materials. However, these materials are only partially accessible, according to legislation and internal structure of the religious institutions. Thus, the use of the materials in Croatia differs between the institutions, and ranges from little access to conventional materials to the possibility of using digitized documents online. International experience and practice, as well as some examples of good practice in Croatia, indicate the need to change the overall approach to the processing and use of this heritage in Croatia.

Key words: archival heritage, religious heritage, religious communities, processing of the materials, use of materials, digitization

Introduction
During the history, the religious institutions played an important role, far more than today, having an impact on the culture, politics, economics, waging the wars, bringing peace and reconstruction. Therefore, European historiography is also based on the archive documents of religious communities, either in their written form or relics of people and events during certain history period.

Croatian regions or Croatians countries were strongly affected by different social, ethnical, political and religious factors and their influences. Speaking about religion, we could see that since the Christianity had spread in this part of Western Europe, that it has been strongly influenced by Catholic Church. However, it must be pointed out that Croatian Catholic identity had been exposed to other influences during the history, so that we can talk about multiethnic, multicultural and multireligious aspects of Croatian identity. It is a combination of different cultural identities, cohesion and assimilation, but at the same time it preserved its dominating and authentic features.[1]

The most important Religious Communities contributing to archival Heritage in Croatia

The existence, presence and the importance of some religious communities was not the same through the history and today in certain historical regions of Croatia. So we can see that northern part of Croatia has had far more different religious communities than the southern part. Historically speaking, Croatia was a Catholic area in Middle Ages, and only in the Modern history we could see the foundation and the dispersion of other religious communities due to the political, economic, social and cultural influences.

Catholic Church[2] based on Roman tradition (Roman Catholic Church) has been present in Croatia since early Middle Ages till today and has been organized in different church hierarchy subdivisions, numerous religious, cultural educational and scientific institutions. Therefore its archival heritage, especially the archives of Catholic Church of Roman tradition is widely spread and present in Croatia. Nevertheless, in spite of the strong influence of Catholic Church, there were other religious communities which spread in Croatia.

In different parts of Croatia throughout the history, within the Catholic and Orthodox context, Eastern Christianity Byzantine tradition got developed as well. Today it is Križevačka eparchy, with its Catholic followers, its institutions, Eastern tradition and regulations, meaning that there is rather big Greek Catholic archival heritage. When we talk about Orthodox religious influence and the influence of its institutions, there is also archival heritage of Serbian Orthodox Church in Croatia.

---

[1] Compare Mohorovičić, Andre. Hrvatska: kulturno-povijesni identitet. // *Hrvatska i europa: kultura, zanost i umjetnost*. Svezak I. Zagreb : HAZU; AGM, 1997., str. 38.

[2] In this paper there are two definitions used Catholic and Roman Catholic. Catholic Church describes all churches of Western and Eastern tradition, liturgy and regulations which are united under the Roman Bishop, Pope. Roman Catholic Church has Western (Latin) tradition, liturgy and regulation. Numerous European, African, Asian churches under the Roman Bishop practice different Eastern traditions, liturgy and regulation. They are also Catholic, but not Roman Catholic. However in Croatia, Catholic Church of Eastern (Byzantine) tradition is called Greek Catholic church.

In 16 and 17 century under the influence of European Reformation, we could see its spreading in Croatia, leaving its archival heritage. The impact of Reformation through the history shows many changes resulting in rather polyvalent churches and religious communities that can be classified under the Reformation heritage. The multipurpose of Reformation churches has strongly influenced their archival heritage.

Judaism and its national group living in Croatia have experienced the same tragic destiny like in other parts of Christian Europe. Their archival heritage shows the permanent exile and an endless fight to preserve their national and religious identity. The documents archived in Croatia, especially since mid 18 century till the onset of II. World War, show their active role in the society and economics within the communities they lived in.

Islam had a negative connotation in Croatia due to the Turkish conquest in 15 century and their rule in 16 and 17 century. It must be mentioned that the Turks occupied some other Balkan regions till the second half of 19 century. Once when the Austrian Hungarian Empire conquered Bosnia and Herzegovina, Islamic community got recognized and from that moment on with the help of the existing documents, we can follow *in situ* the influence of Islam in north Croatia regions and its archival heritage.

The contemporary functioning of the religious communities in Croatia is prescribed by the *Law on legal status of religious communities*[3]. The registered religious communities in Croatia, not only the ones long and traditionally established, but also the new ones, are registered or are in the process of being registered in the *Religious community register of Croatia*[4], which is kept by the Ministry of Administration according to the passed law. For the institutions of Catholic Church there is special *Register of institutions of Catholic Church in Croatia*[5]. Both registers are available *online* on the web pages of Croatian Ministry of Administration.[6]

The religious communities officially listed in the Register were described in book with a systematic overlook on existing data about the religious communities in Croatia through the history and today.[7] In this book the religious communities are discussed from the perspective of the religious sociology, but in order to study, understand, preserve and use archival heritage,

---

[3] Com. Zakon o pravnom položaju vjerskih zajednica. *Narodne novine* 83/02.

[4] Evidencija vjerskih zajednica u Republici Hrvatskoj. URL: http://www.appluprava.hr/ RegistarVjerskihZajednica/ (2011-9-9).

[5] Evidencija pravnih osoba Katoličke Crkve u Republici Hrvatskoj. URL: http://www.appluprava. hr/PravneOsobeKatolickeCrkve/ (2011-9-9).

[6] Ministarstvo uprave RH. URL: http://www.uprava.hr/ (2011-9-9).

[7] Com. Marinović Bobinac, Ankica; Marinović Jerolimov, Dinka. Vjerske zajednice u Hrvatskoj. Zagreb : Prometej; Udruga za vjersku slobodu u RH, 2008.

some other methods should be introduced. It is necessary to know how the religious communities operated within some system, network, relations among offices, governments, municipalities in certain region. It can be done only with the use of high quality literature data, together with the regulations of certain religious community and nowadays with the help of online/web data, what can be of an enormous benefit for the extensive research.

The thorough data investigation of the religious community founders in Croatia for the purpose of archival public was conducted by archivist and historian Josip Buturac in1970 and it can be seen as a basis of the history archives of religious communities. In this book we could find history data about the governing of religious communities in Croatia from the very beginning till 20 century, the history of religious institutions, their hierarchy and the history overlook of the dominant religious communities in Croatia, for e.g.: Roman Catholic and Greek Catholic church, with the emphasis on catholic monks; Serbian Orthodox church; Evangelical church, Reformed church and Jewish religious community.[8] Nevertheless, there were other publications published by some religious communities, which could help with the research of their institutions and their history. Recently, they are available on web/online documents, enabling a fast access to summarized information about the history and the functioning of the religious institutions, either in the past or today.

The author of this paper has studied the different regulations of religious communities in his thesis[9] and talked about them on the seminar "The preservation of religious community documentation" in Croatian national archive in 2010.[10] The thesis presents the typology and hierarchical archive levels of Croatian religious communities[11], typology and the importance of the preserved archives content[12], including the most important data resources and their application[13].

---

[8] Com. Buturac, Josip. Povijest uprave vjerskih organizacija u Hrvatskoj od X-XX stoljeća, Zagreb : Arhiv Hrvatske, 1970., str. 7-65.

[9] Com. Kušen, Dražen. Arhivi vjerskih zajednica u sjevernoj Hrvatskoj. Razvoj, tipologija, sadržajni značaj. Doktorski rad na Filozofskom fakultetu Sveučilišta u Zagrebu, Zagreb, 2011., str. 62-194.

[10] Com. Kušen, Dražen. Arhivi u propisima vjerskih zajednica – od pismohrane do ustanove. // *Zaštita dokumentacije vjerskih zajednica. Seminar. Hrvatski državni arhiv, Zagreb, 19. i 20. travnja 2010. (sažeci izlaganja)*. Zagreb : Hrvatski državni arhiv, 2010., str. 23-34.

[11] Com. Kušen, Dražen. Arhivi vjerskih zajednica u sjevernoj Hrvatskoj, str. 62-194.

[12] Idem, str. 195-233.

[13] Idem, str. 234-240.

Archival heritage of the Religious Communities in Croatia – Register and Access

According to the *Law on archive material and archives*, the archive material of the religious communities is seen as the private archive material in Croatia, so it is regulated with the certain protocols within the Law.[14] Their own regulations regarding the archive material, it's saving, preserving and use has only Catholic Church in Croatia. The other religious communities would have these regulations only regarding the church administration, keeping the registers of births, marriages and deaths, and their finances, but it is must be pointed out that some of them have started issuing their own archive regulations.

The history context confirms that the religious archives are together with their development, typology and the content an, essential part of the archival heritage, dating from early Croatian history till nowadays.[15] The heritage records are mostly kept by the religious communities, but the approach to these data differs depending on the institution. Speaking about the public, the most important archival heritage of the religious communities is recorded in *Register of archive funds and collections of Republic of Croatia*, which was published in 2006.[16] The same data published in the register are also available via ARHiNET internet system.[17] The online access of these documents and its importance and potential was presented by Vlatka Lemić and Tomislav Čepulić on the seminar "The preservation of the religious communities documents".[18]

This archival heritage is pretty dispersed in two different contexts. On one side they are the holdings and collections of archival heritage of religious communities, which are preserved in public archive institutions (Croatian State Archives, its branch offices/archives and some other archives or institutions) as their own archive materials. On the other side they are holdings and collections which are preserved within the religious institutions and archives as their own materials.

---

[14] Com. Zakon o arhivskom gradivu i arhivima. *Narodne novine* 105/97.

[15] Com. Kolanović, Josip. Arhivi u Socijalističkoj Republici Hrvatskoj. // Arhivski fondovi i zbirke u SFRJ; SR Hrvatska. Beograd : Savez Arhivskih radnika Jugoslavije, 1984., str. 11-37.

[16] Pregled arhivskih fondova i zbirki Republike Hrvatske. Zagreb : Hrvatski državni arhiv, 2006.

[17] ARHiNET. // URL: http://arhinet.arhiv.hr/. See also: Lemić, Vlatka; Čabrajić, Hrvoje. Managing and Presenting Digital Content in the ARHiNET System. // *The future of information sciences: INFuture 2009 - Digital resources and knowledge sharing : [2nd international conference, Zagreb, 4-6 November 2009]*; edited by Hrvoje Stančić ... [et al.]. Zagreb : Department of Information Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, 2009., str. 95-104.

[18] Com. Lemić, Vlatka; Ćepulić, Tomislav. Arhivska služba u RH i vjerske zajednice – korištenje dokumentacije vjerskih zajednica. // *Zaštita dokumentacije vjerskih zajednica,* str. 41-42.

In *Register of archive fonds and collections of Republic of Croatia*[19], it was possible to present all registered archive material of religious communities in both contexts, meaning not only the material saved in the archives, but also in some religious institutions. Besides the data which got published as the result of the cooperation between Croatian state archive and the religious communities, most religious communities do not have any published registers of their own archives.

The archives in the religious institutions are not only the records about the history of the very institutions, but also other holdings and collections, gathered, donated or deposited during the history.[20] Most of the documents have never been stored or preserved in the public archives. The most important documents kept in religious archives are: notarial documents, register of births, marriages and deaths, records of Canon visitations (inspections) and *Status animarum*.[21]

The public has a limited access to religious archives due to their own regulations and the organization within the religious communities. We could say that possibility to access some of these archives varies in different institutions. There are some archives which are easy to access, including via internet, and the other ones with limited or no access to rather at all conventional documents. The basic information about the preserved documents of the religious communities can be used online via ARHiNET internet system, and some religious communities, or Croatian State Archives, have their documents digitised, including their most important materials, like the registers of births, marriage and deaths.

The registers of births, marriage and death have been carefully preserved for centuries either in the public archives or in the religious archives. The access to these registers is legally prescribed by the state law or international conventions[22], consequently the access to these registers is limited. A lot of religious registers (older than 100 years) are preserved in the public archives, since they were taken from the religious archives after the II. World War. Therefore the public archives preserve them, protect their screening and usage, like they do with all other funds in Croatian State Archives and other public archives in Croatia.[23] The religious registers are extremely important for the

---

[19] Com. Pregled arhivskih fondova i zbirki Republike Hrvatske.

[20] Com. Pavliček, Vida. Pregled arhivskih fondova i zbirki Republike Hrvatske. Rad na prikupljanju i obradi podataka. // Arhivski vjesnik, 50 (2007), str. 13-14.

[21] Com. Kušen, Dražen. Arhivi vjerskih zajednica u sjevernoj Hrvatskoj, str. 234-240.

[22] Com. Zaštita osobnih podataka i dostupnost informacija : Preporuke Vijeća Europe. URL: http://www.arhiv.hr/hr/pdf/Zastita%20osobnih%20podataka.pdf (2011-6-28).

[23] Com. Baričević, Zvonko. Zaštita kulturnih dobara – onemogućavanje nezakonitog prometa kulturnim dobrima. Crkveni arhivi u središnjoj Europi – tradicija, funkcije, suradnja, izazovi. // *Zaštita dokumentacije vjerskih zajednica*, str. 15-21.

genealogical and demographic researches, so they can provide us with the essential information about the population living in one region in a specific period. There are also scientific papers based on the materials coming from registers and the papers explaining the register history value and the importance of the documents contained in the registers. When we talk about the demographic research, registers are of a great importance for the religious community or other social demographic studies, philological research, medical heritage studies, as well as the closest source of the information for the research or determination of the genealogy.

In most other countries the old registers of births, marriage and death are dispersed and kept by the various institutions (church/religious offices or public/national register offices) and therefore are difficult to have an easy access for a systematic research. There are examples in the world where the registers in the bishop's archive are deliberately collected, preserved and used. The very illustrative example of how the registers can be used in the research, is in Catholic diocese in Passau, Germany, which launched the project of creating database of Passau population before 1900.[24] Working on the project, they used all the registers of Passau parishes, which were kept in bishop's archives (what is typical for Bavaria, but not in other parts of Germany). The first project phase consisted of data entry of all registers and creation of the network and the second phase would be digital filing of all sources and connecting records with central database. The purpose of the project is not only to enable an easy access to data, but also to prevent further use of the original source and its damage.[25] The project was rather demanding, because it meant to information and IT process and connect data of 6000 registers with more than 4 million data to be stored, so that the necessary information and all other linked pages could be easily identified and browsed. After 10 years of intensive work of the Passau Bishop archive and IT experts the result was outstanding. The database content was immediately used in Alzeheimer disease research and it was of a great help when the family relations or genealogy were studied, using a special application which automatically browsed the data.[26] This successful project is an example of how the register records can be systematically collected with the protective scanning, how the data can be prepared for IT processing and how the use of the

---

[24] Com. Wurster, Herbert W. The Historical Population Data Base of the Catholic Diocese of Passau, Germany, and its Use for Genetic Research. // URL: http://www.wien2004.ica.org/imagesUpload/pres_283_WURSTER_SKR03.pdf?PHPSESSID=7230ed24fb7c9533364e768725 582fb4 (2009-09-18).

[25] Com. Workshop Herbert W. Wurster. Usp. URL: http://www.ddb.umu.se/aktuellt/workshop dec2003/herbertwwurster.html (2009-9-24).

[26] Com. Project desscription for GenDB. Usp. URL: http://www.forwiss.uni-passau.de/en/ projectsingle/19/main.html (2009-9-24).

registers can stimulate the cooperation of religious communities and the public archives.

The audio visual materials, not only the new ones, but also the very first ones of this kind can be also found in the various religious archives.[27] However the new electronic data is of great concern, because its existence is not recorded in the national archives, but has been stored in the religious archive for decades.[28] This material should be processed, applying the international and national methods. The Department of archive and documentation within the Department of Information Science at Faculty of Humanities and Social Sciences in Zagreb in its education of archivists and other experts has been emphasizing strongly the conservation of electronic data. This conference has had it as its goal since 2007.[29] The gained and shared experiences, study results in information science and plied in heritage institutions in public offices, should help us shape the right policy toward electronic data, other sources and private religious institutions. The outcome of this modern policy should be wide usage of modern technology to protect and enable easier access to archives, including the ones in religious institutions. The modern technology should be particulary used in the case of very important materials and if there are more inquiries. The good examples are ICARUS[30] and Metis programs[31].

Some examples of good practice in Croatia

The outstanding example of a good cooperation between the religious institution and the national archive is one hundred year old cooperation between Archbishop diocese in Zagreb and Croatian State Archives. The Archbishop archive is the most important archive of Catholic church in Croatia, because it not only preserves the records of its own activities and of its offices, but also is in the possession of the materials of many other church institutions within Archbishop diocese. The whole archive holdings dating from very early ages till 1945, respecting the agreement of both parties' obligations and rights is deposed in Croatian State Archives. All the holdings (mostly manuscripts from Middle ages and contemporary) together with records of Bishop library of Zagreb Archbishop diocese (from Middle Ages, mostly 14 century) are

---

[27] Com. Lhotka, Carmen. Hrvatska kinoteka – filmsko gradivo kao povijesni izvor. // *Zaštita dokumentacije vjerskih zajednica*, str. 41-42.

[28] The author of this work has their own insight into the more parochial and monastic archives of the Catholic Church in Croatia, and into the archives of the Jewish community of Osijek.

[29] Com. INFuture 2011. // URL: http://infoz.ffzg.hr/infuture/ (2011-6-27).

[30] ICARUS. // URL: http://www.icar-us.eu (2010-06-20).

[31] Com. Colombo, Massimo. Metis: Italian Technology for the worldwide Cultural Heritage. *Atlanti*, Vol. 17, N. 1-2, Trieste 2007, pp. 299-300. // URL: http://www.iias-trieste-maribor.eu/ fileadmin/atti/2007/Colombo.pdf (2007-10-12).

preserved in Croatian State Archives at the Department of Zagreb Archbishop diocese.[32] We can say that of all religious institutions of Catholic church in Croatia, Zagreb Archbishop diocese has solved best the problem of archive preservation, the filing of the documents and their use. The preserved materials are precious source for the research of North Croatia history. The History Association of Archbishop diocese of Zagreb "Tkalčić" is very active in systematic research, publishing and preservation of the heritage, as well as in the holding presentation to the public (also via internet).[33]

Talking about Jewish communities in Croatia, we must mention an extremely successful project, conducted by the Research and documenting center "Cendo". The center got organized on the international level in 2000 in Jewish community of Zagreb and collects materials, documents and testimonies of the Jews living in Croatia. It also keeps the archive, publishes the catalogues, organizes data base for its users, studies all history segments of Jews in Croatia and publishes the research results and gathers data.[34]

Preservation of Religious Communities' Archives and their Accessibility on the international Level

International Council on Archives (ICA)[35] has a special department "Section of Archives of Churches and Religious Denominations" (SKR), which got founded in 1995 in Prag and according to its regulations, works on accessibility and the protection of cultural resources which should promote the understanding of the world religions. ICA/SKR has its own publication "Newsletter of ICA Section of Archives of Churches and Religious Denominations".[36]

UNESCO on line material could provide also the links of the religious communities archives all over the world, as well as web pages of worldly, regional or national religious associations, organizations or specialized associations which take care of religious heritage archive.[37] We could mention

---

[32] Com. Razum, Stjepan. Biskupijski, kaptolski i samostanski arhivi. Nadbiskupijski arhiv Zagreb i Metropolitanska knjižnica Zagreb. // *Zaštita dokumentacije vjerskih zajednica*, str. 35-36. See also: Državni arhiv u Zagrebu http://www.daz.hr/arhol/prilozi_arhivi.htm (2010-5-3)

[33] Com. Tkalčić : Društvo za povjesnicu Zagrebačke nadbiskupije. // URL: http://www.tkalcic.hr (2010-5-28).

[34] CENDO. // URL: http://www.zoz.hr/home.php?content=content&term=39&key=26&key1=39 (2010-6-14).

[35] ICA : International Council on Archives. // URL: http://www.ica.org (2010-5-26).

[36] ICA/SKR : Section of Archives of Churches and Religious Denominations. // URL: http://www.ica.org/en/node/30820#ica-infobox-2 (2010-5-26).

[37] Com. UNESCO Archives Portal. // URL: http://www.unesco-ci.org/cgi-bin/portals/archives/page.cgi?d=1 (2010-2-8).

only few archive associations of religious communities like the ones in Italy[38], Germany (evangelical[39] and catholic[40]), Austria (diocese[41] and monastery[42]) and Hungary (interreligious)[43]. All these associations via their on line documents and web pages can supply a lot of information about their work, about the work of the religious communities, the community they function in, their own archives and their holdings.

Conclusion

The condition, the objective limitation, but at the same time the great potentials of the archives of the religious communities in Croatia raise the questions of an necessary preservation of these archives and their accessibility for the research purposes. If we take international experience, Croatia should also try to have more and more on line access to the archives. Nevertheless, the access to electronic (digitised) data of religious archives is not easy to achieve in Croatia, due to human and financial limitations. In the future the cooperation of the religious institutions with the archive experts from the public archives would improve the education and training of the people working in the religious archives, resulting in better e-access to their materials. However, in order to achieve the goal of an easy access to archive material, it is necessary to improve the cooperation between private religious institutions and public archives, what would have as a result better preservation, valorization, conservation and the use of archive materials which are kept in the archives of the religious communities.

References

Arbeitsgemeinschaft der Ordensarchive (AGOA). // URL: http://www.orden.de/index_oa.php (2010-2-8).

Arbeitsgemeinschaft der Ordensarchive Österreichs. // http://www.ordensarchive.at/index.php (2010-2-7).

ARHiNET. // URL: http://arhinet.arhiv.hr/.

Baričević, Zvonko. Zaštita kulturnih dobara – onemogućavanje nezakonitog prometa kulturnim dobrima. Crkveni arhivi u središnjoj Europi – tradicija, funkcije, suradnja, izazovi. // *Zaštita*

---

[38] Com. Archiva ecclesiae. // URL: http://www.archivaecclesiae.org/index.php?option=com_frontpage&Itemid=1 (2010-2-8).

[39] Com. Verband kirchlicher Archive. // URL: http://www.ekd.de/archive/index.htm (2010-2-7).

[40] Com. Arbeitsgemeinschaft der Ordensarchive (AGOA). // URL: http://www.orden.de/index_oa.php (2010-2-8).

[41] Com. Das Portal zum Archivwesen der römisch-katholischen Kirche in Österreich. // URL: http://www.kirchenarchive.at (2010-2-6).

[42] Com. Arbeitsgemeinschaft der Ordensarchive Österreichs. // http://www.ordensarchive.at/index.php (2010-2-7).

[43] Com. MELTA. // URL: http://archivum.piar.hu/melte/ (2011-6-25).

dokumentacije vjerskih zajednica. Seminar. Hrvatski državni arhiv, Zagreb, 19. i 20. travnja 2010. (sažeci izlaganja). Zagreb : Hrvatski državni arhiv, 2010., str. 15-21.

Buturac, Josip. Povijest uprave vjerskih organizacija u Hrvatskoj od X-XX stoljeća, Zagreb : Arhiv Hrvatske, 1970.

CENDO. // URL: http://www.zoz.hr/home.php?content=content&term=39&key=26&key1=39 (2010-6-14).

Colombo, Massimo. Metis: Italian Technology for the worldwide Cultural Heritage. *Atlanti*, Vol. 17, N. 1-2, Trieste 2007, pp. 299-300. // URL: http://www.iias-trieste-maribor.eu/fileadmin/atti/2007/Colombo.pdf (2007-10-12)

Das Portal zum Archivwesen der römisch-katolischen Kirche in Österreich. // URL: http://www.kirchenarchive.at (2010-2-6)

Državni arhiv u Zagrebu http://www.daz.hr/arhol/prilozi_arhivi.htm (2010-5-3)

Evidencija pravnih osoba Katoličke Crkve u Republici Hrvatskoj. URL: http://www.appluprava.hr/PravneOsobeKatolickeCrkve/ (2011-9-9).

Evidencija vjerskih zajednica u Republici Hrvatskoj. URL: http://www.appluprava.hr/Registar VjerskihZajednica/ (2011-9-9).

ICA : International Council on Archives. // URL: http://www.ica.org (2010-5-26).

ICA/SKR : Section of Archives of Churches and Religious Denominations. // URL: http://www.ica.org/en/node/30820#ica-infobox-2 (2010-5-26).

ICARUS. // URL: http://www.icar-us.eu (2010-06-20).

INFuture 2011. // URL: http://infoz.ffzg.hr/infuture/ (2011-6-27).

Kolanović, Josip. Arhivi u Socijalističkoj Republici Hrvatskoj. // *Arhivski fondovi i zbirke u SFRJ*; SR Hrvatska. Beograd : Savez Arhivskih radnika Jugoslavije, 1984., str. 11-37.

Kušen, Dražen. Arhivi u propisima vjerskih zajednica – od pismohrane do ustanove. // *Zaštita dokumentacije vjerskih zajednica. Seminar. Hrvatski državni arhiv, Zagreb, 19. i 20. travnja 2010. (sažeci izlaganja)*. Zagreb : Hrvatski državni arhiv, 2010., str. 23-34.

Kušen, Dražen. Arhivi vjerskih zajednica u sjevernoj Hrvatskoj. Razvoj, tipologija, sadržajni značaj. Doktorski rad na Filozofskom fakultetu Sveučilišta u Zagrebu, Zagreb, 2011.

Lemić, Vlatka; Ćepulić, Tomislav. Arhivska služba u RH i vjerske zajednice – korištenje dokumentacije vjerskih zajednica. // *Zaštita dokumentacije vjerskih zajednica. Seminar. Hrvatski državni arhiv, Zagreb, 19. i 20. travnja 2010. (sažeci izlaganja)*. Zagreb : Hrvatski državni arhiv, 2010.*, str. 41-42.

Lemić, Vlatka/ Čabrajić, Hrvoje. Managing and Presenting Digital Content in the ARHiNET System. // *The future of information sciences: INFuture 2009 - Digital resources and knowledge sharing : [2nd international conference, Zagreb, 4-6 November 2009]*; edited by Hrvoje Stančić ... [et al.]. Zagreb : Department of Information Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, 2009., str. 95-104.

Lhotka, Carmen. Hrvatska kinoteka – filmsko gradivo kao povijesni izvor. // *Zaštita dokumentacije vjerskih zajednica. Seminar. Hrvatski državni arhiv, Zagreb, 19. i 20. travnja 2010. (sažeci izlaganja)*. Zagreb : Hrvatski državni arhiv, 2010., str. 41-42.

Marinović Bobinac, Ankica; Marinović Jerolimov, Dinka. Vjerske zajednice u Hrvatskoj. Zagreb : Prometej; Udruga za vjersku slobodu u RH, 2008.

MELTA. // URL: http://archivum.piar.hu/melte/ (2011-6-25).

Ministarstvo uprave RH. URL: http://www.uprava.hr/ (2011-9-9).

Mohorovičić, Andre. Hrvatska: kulturno-povijesni identitet. // *Hrvatska i europa: kultura, zanost i umjetnost*. Svezak I. Zagreb : HAZU; AGM, 1997., str. 38.

Pavliček, Vida. Pregled arhivskih fondova i zbirki Republike Hrvatske. Rad na prikupljanju i obradi podataka. // *Arhivski vjesnik*, 50 (2007), str. 13-14.

Pregled arhivskih fondova i zbirki Republike Hrvatske. Zagreb : Hrvatski državni arhiv, 2006.

Project description for GenDB. Usp. URL: http://www.forwiss.uni-passau.de/en/projectsingle/19/main.html (2009-9-24).

Razum, Stjepan. Biskupijski, kaptolski i samostanski arhivi. Nadbiskupijski arhiv Zagreb i Metropolitanska knjižnica Zagreb. // *Zaštita dokumentacije vjerskih zajednica. Seminar. Hrvatski državni arhiv, Zagreb, 19. i 20. travnja 2010. (sažeci izlaganja)*. Zagreb : Hrvatski državni arhiv, 2010., str. 35-36.

The future of information sciences: INFuture 2009 - Digital resources and knowledge sharing : [2nd international conference, Zagreb, 4-6 November 2009]; edited by Hrvoje Stančić ... [et al.]. Zagreb : Department of Information Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, 2009.

Tkalčić : Društvo za povjesnicu Zagrebačke nadbiskupije. // URL: http://www.tkalcic.hr (2010-5-28).

UNESCO Archives Portal. // URL: http://www.unesco-ci.org/cgi-bin/portals/archives/page.cgi?d=1 (2010-2-8) Archiva Ecclesiae. URL: http://www.archivaecclesiae.org/index.php?option=com_frontpage&Itemid=1 (2010-2-8)

Verband kirchlicher Archive. // URL: http://www.ekd.de/archive/index.htm (2010-2-7).

Workshop Herbert W. Wurster. Usp. URL: http://www.ddb.umu.se/aktuellt/workshopdec2003/herbertwwurster.html (2009-9-24).

Wurster, Herbert W. The Historical Population Data Base of the Catholic Diocese of Passau, Germany, and its Use for Genetic Research. // URL: http://www.wien2004.ica.org/images Upload/pres_283_WURSTER_SKR03.pdf?PHPSESSID=7230ed24fb7c9533364e768725582 fb4 (2009-09-18).

Zakon o arhivskom gradivu i arhivima. *Narodne novine 105/97.*

Zakon o pravnom položaju vjerskih zajednica. *Narodne novine 83/02.*

Zaštita dokumentacije vjerskih zajednica. Seminar. Hrvatski državni arhiv, Zagreb, 19. i 20. travnja 2010. (sažeci izlaganja). Zagreb : Hrvatski državni arhiv, 2010.

Zaštita osobnih podataka i dostupnost informacija : Preporuke Vijeća Europe. URL: http://www.arhiv.hr/hr/pdf/Zastita%20osobnih%20podataka.pdf (2011-6-28).

Presence of Institutional Culture in Daily Newspapers – Analysis of Printed Editions of *Jutarnji list* and *Večernji list* in 2009

Dunja Majstorović

Postgraduate doctoral study in Information and Communication Sciences,
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, 10000 Zagreb, Croatia
dunja.majstorovic@gmail.com

Summary

This paper focuses on the analysis of the relationship between culture and media through presence of institutional culture in daily newspapers – Jutarnji list (JL) and Večernji list (VL). The aim was to get a quantitative indicator of the presence of institutional culture in two analyzed newspapers. In the course of research the following hypotheses were made: 1) Despite the existence of a culture column in the analyzed newspapers, the amount of cultural content is relatively small, 2) There is no significant difference in size of the culture column and the number of published texts on culture in the two analyzed newspapers, 3) The largest number of published texts is in the form of news. The timeframe of the study was one month (from 23 November 2009 to 20 December 2009) and the used method of research was content analysis. Found texts related to culture were sorted by type of text into the following categories: news, short articles, articles and interviews.
The total number of found texts related to culture in JL was 315, and 355 in VL. The results of the study have confirmed the hypothesis related to low presence of culture in daily newspapers (first and third hypothesis) since most of the published texts were in the form of news and short articles (in JL they accounted for 79% and in VL as much as 87% of published texts). The second hypothesis was not fully confirmed. In fact, a significant difference in the size of the JL and VL culture columns was not established (both amount to approximately 2 pages), but there is a difference when it comes to the comparison of texts by type.

Key words: presence of institutional culture in daily newspapers, content analysis, Jutarnji list, Večernji list

Introduction

The relationship between culture and the media is very complex. However, in order to be able to write about their relationship, it is necessary to first define

the two fundamental concepts. Denis McQuail has defined mass media as means of communication that operate on a large scale, reaching and involving, in greater or lesser extent, almost the entire society. The mass media, therefore, include: newspapers, magazines, film, radio, television and recorded music.[1]

Unlike the media, culture is a term that is much harder to define. It originates from the Latin term "cultus" which means cultivation, care, upbringing, farming, education, worship.[2] The practical use of the term culture began in Germany in the 18th century and in the 19th century the term was given a scientific determination by E.B. Taylor in his Primitive cultures (1871). Taylor defined culture as a "complex whole which includes knowledge, belief, art, morals, law, customs and any other capabilities and habits acquired by man as a member of society."[3].

There are many different definitions of culture and in everyday speech culture is often equated with civilization or arts.

Precisely to avoid the overly broad conception of culture, which, inter alia, is defined as "the totality of its various forms"[4] or even as "...a process of humanization, characterized by a collective effort to preserve human life, to stop the struggle for existence...to...develop the spiritual capacity of people to reduce and sublimate aggression, violence and misery"[5], within which it is possible to include even the media, this paper will focus exclusively on the presence of institutional culture in the newspapers.

The media is the most important informer on all cultural events and its news, reviews, critics and other forms of news reporting inform the public about the recent cultural production and thus critically evaluate it[6]. It is obvious that the media and culture are interconnected but their positions are unequal. The culture columns, although present in most of the media, are minimized, while on the other hand, the question is raised whether culture can even survive without the media? This research was conducted to determine the position of institutional culture (in culture columns but also in other newspaper columns) in printed editions of two national daily newspapers – Jutarnji list and Večernji list (sample of the month from 23 November 2009 to 20 December 2009) as well as

---

[1] McQuail, Denis. McQuail's mass communication theory. London : Sage Publications, 2000; p. 4.

[2] Kale, Eduard. Uvod u znanost o kulturi. Zagreb : Školska knjige, 1977.; p. 38.

[3] Taylor, Edward. B. Primitive culture (taken from: Moore, Jerry D. Uvod u antropologiju : Teorije i teoretičari kulture. Zagreb: Naklada Jesenski i Turk, 2002.; p. 21-22.

[4] Definition by Alfred Kroeber (taken from: Kale, Eduard. Uvod u znanost o kulturi. Zagreb : Školska knjige, 1977.; p. 40.)

[5] Definition by Herbert Marcuse (taken from: Sadžakov, Slobodan. Marcuseovo shvaćanje kulture. // Filozofska istraživanja. 28 (2003), 6; p. 118.)

[6] Malović, Stjepan. Mediji i društvo. Zagreb: ICEJ i Sveučilišna knjižara, 2007.; p. 65.

to confirm the inferior position of culture, which is minimized in relation to entertainment and sports[7].

Research of the presence of institutional culture in Jutarnji list and Večernji list

The research of presence of institutional culture was conducted in printed editions of two daily newspapers – Jutarnji list (JL) and Večernji list (VL). In selecting the newspapers for the research, attention was paid to national distribution and the number of readers. In Croatia there are four national daily newspapers: 24 sata, Jutarnji list, Večernji list and Vjesnik. The newspaper 24 sata was excluded from the research because it is a tabloid that does not have a culture column and Vjesnik, although it has a renowned culture column, was excluded due to the small number of readers. Therefore, it was decided that the research should be conducted on Jutarnji list and Večernji list because of their similarities: both newspapers are established in the form of semi-tabloids which are characterized by short texts and distinctive photographs along with regular columns of comments and views[8].

Also, it was decided that the research will be carried out on printed editions of JL and VL, because the contents of web editions constantly change (many times during the day).

Research methodology

The aim of the research was to get a quantitative indicator of the presence of institutional culture (classified in categories) in two daily newspapers.

In the course of the research three hypotheses were made:

1. Despite the existence of a culture column in the analyzed newspapers, the amount of cultural content is relatively small.
2. There is no significant difference in size of the culture column and the number of published texts on institutional culture in the two analyzed newspapers.
3. The largest number of published texts is in the form of news – only brief information about a particular cultural event.

The timeframe of the content analysis of JL and VL was one month (from November 23rd 2009 to December 20th 2009).

The found texts related to institutional culture, within the culture column, but also featured in other newspaper columns[9], were sorted:

---

[7] Malović, Stjepan. Odnos masovnih medija i kulture : globalni trendovi i domaća lutanja. // Kolo, 15 (2005), 2; p. 249.

[8] Dragojević, Sanjin; Kanižaj, Igor; Žebec, Ivana. Europska Unija u hrvatskim dnevnim novinama : avis – odgoda pregovora. // Politička misao. 43 (2006), 3; p. 138.

[9] The research was primarily focused on culture columns, but since the texts on institutional culture can sometimes be found in other columns, like Guide or Spectacles, these texts were also

by type of text into the following categories:
- news – up to 35 lines of text,
- short articles – up to half a page of text,
- articles – over half a page of text,
- interviews;

according to the division of culture[10] into categories:
- books and publishing,
- visual arts,
- music and discography,
- theatre (performing arts),
- cinematography,
- libraries and archives,
- museums and galleries,
- cultural heritage,
- other - festivals, lectures, audience, performance, meetings…

In order to gain insight about the size of the culture column in relation to other columns in the analyzed newspapers JL and VL, the number of pages of each column by days of the week (one week in the investigated period of one month) was recorded in the table. After that, the average number of pages of all columns was calculated.

Structure of the analyzed newspapers and place of culture column

In Croatia there are no recognizable stylistic characteristics of newspapers. Readers are confused by the combination of styles, so in the newspapers that want to write about serious topics we encounter articles that are stylistically close to tabloids.[11] This can be read from the structure of the analyzed newspapers, especially from the structure of columns and their average volume.

Jutarnji list

The owners of JL are Europa Press Holding (EPH) and Westdeutsche Allgemeine Zeitung (WAZ). It was first published in 1998.

JL has 16 columns in the following order: News, Black Chronicle, World, Money (except Fridays and Sundays), Culture, Comments, Sport, In Memoriam, Jutarnji 2 (except Saturdays and Sundays), Spectacles, Guide, Entertainment, Weather, TV program and the last page. Sunday edition is not

---

included. Other texts, such as comments/essays of leading Croatian writers were excluded mainly because they too often write about politics and other subjects different from culture. The other reason was the focus on institutional culture as opposed to culture in general.

[10] With modifications, the division of culture was taken from the official website of the Croatian Ministry of Culture. Available at: http://www.min-kulture.hr (18 April 2011)

[11] Malović, Stjepan. Medijski prijepori. Zagreb : Izvori, 2004.; p. 99-101.

very different from the weekly edition. Apart from the usual columns (except column Money and Jutarnji 2) two other columns appear (Panorama and Plus). The average total number of pages in JL is 46.72. Column News has the highest average number of pages (10.28 or 22%). It is followed by Sport (7.86 pages or 17%) and Jutarnji 2 (5.86 pages or 12%).

The culture column takes an average of two pages or 4% of total pages, and is located in front of column Sport and Spectacle. Other columns in JL amount to average of 45% of the total content of newspapers.

Večernji list

This daily national newspaper was first published in 1959. Since the year 2000 it has been owned by the Styria Medien AG, the Austrian media group.

VL has 10 regular columns (six less than in JL) in the following order: News, Panorama, Business, Open matters, Obituaries, Sports, Entertainment, Culture and stage, Compass (except Fridays and Saturdays), TV program and last pages. Sunday edition is significantly different from the issues during the week: there are no columns – Panorama, Business, Open, Entertainment, Culture and scene, but there are other columns – Spectacles, Junior fun, Food, Garden, Pets, Horoscope and Medical consulting.

The average total number of pages of VL is 47.23. The largest number of pages is dedicated to the News column (an average of 16.14 pages or 34%), then Panorama (8 pages or 17%) and Sport (7 pages, or 15%). The fourth place belongs to the column Culture and stage (3.71 pages or 8%). The culture column in VL is found behind Sports and Entertainment; along with the column stage (the very order of the columns indicates their importance).

It is important to note that within the column Culture and stage, the culture column (cultural events) alone usually amounts to two pages or 4%, which is the same extent as in JL. Other columns in VL amount to an average of 26% of the total content of the newspapers.

Research results

Quantitative indicators of types of texts according to the division of culture are presented in Table 1 (JL) and Table 2 (VL).

Jutarnji list – quantitative indicators

In JL in the course of research a total number of 315 texts were published. The highest amount of published texts was in the form of news – 148 or 47% of total texts. The number of short articles was 100 (32% of total texts) and the number of articles was 60 (19% of total texts). There were only 7 published interviews (which makes just 2% of total texts).

As much as 79% of published texts on cultural events were in the form of news and short articles. These data confirm the first hypothesis of relatively small amount of cultural content in the analyzed newspapers. Even though the number

of published news and short articles is relevant (248 texts), it is impossible to overcome the fact that those are brief pieces of information that can satisfy the quantity but not the quality of the contents. The third hypothesis, which stated that the largest number of the published texts were in the form of the news, was also confirmed.

Table 1: Jutarnji list – quantitative indicators of types of texts according to the division of culture

DIVISION OF CULTURE BY CATEGORIES	TYPE OF TEXT				
	News	Short article	Article	Interview	Total
Books and publishing	29	11	13	3	56
Visual arts	13	4	7	1	25
Music and discography	18	25	18	2	63
Theatre (performing arts)	29	25	3	0	57
Cinematography	15	18	8	0	41
Library and archives	0	0	0	0	0
Museums and galleries	24	17	9	1	51
Cultural Heritage	3	0	1	0	4
Other	17	0	1	0	18
Total	148	100	60	7	315

If we consider the structure of the published texts in JL, according to the culture division categories, the results are as follows:

- Books and publishing – 56 texts (or 18% of the published texts). The structure according to the type of texts: 29 news (or 52%), 13 articles (23%), 11 short articles (or 20%) and 3 interviews (or 5%).
- Visual arts – 25 texts (8% of the published texts). The structure according to the type of text: 13 news (or 52%), 7 articles (or 28%), 4 short articles (or 16%) and 1 interview (or 4%).
- Music and discography – 63 texts (20% of the published texts). The structure according to the type of text: 25 short articles (or 40%), 18 news (28.5%), 18 articles (or 28.5%) and 2 interviews (or 3%).
- Theatre (performing arts) – 57 texts (18% of the published texts). The structure according to the type of text: 29 news (or 51%), 25 short articles (or 44%), and 3 articles (or 5%).
- Cinematography – 41 texts (13% of the published texts). The structure according to the type of text: 18 short articles (or 44%), 15 news (or 37%) and 8 articles (or 19%).

- Museums and galleries – 51 texts (16% of the published texts). The structure according to the type of text: 24 news (or 47%), 17 short articles (or 33%), 9 articles (18%) and 1 interview (or 2%).
- Cultural heritage – 4 texts (or 1.3% of the published texts).
- Other – 18 texts (5.7% of the published texts).

It is interesting that in the category of libraries and archives, in the analyzed period, not a single text was published.

Večernji list – quantitative indicators
In VL, in the analyzed period, a total number of 355 texts were published. The highest amount of published texts was in the form of news (273 or 77% of total texts). There were 37 short articles (or 10%), 35 articles (or approximately 10%) and 10 interviews (or 3% of total texts).

As much as 87% of published texts on cultural events were in the form of news and short articles, and only13% of the totally published texts in VL, by size, are more serious articles. These data, as well as in the case of JL, confirm the first and the third hypotheses of the research.

Table 2: Večernji list – quantitative indicators of types of texts according to the division of culture

DIVISION OF CULTURE BY CATEGORIES	TYPE OF TEXT				
	News	Short article	Article	Interview	Total
Books and publishing	45	13	6	2	66
Visual arts	7	0	6	0	13
Music and discography	82	8	7	6	103
Theatre (performing arts)	46	1	4	1	52
Cinematography	40	4	4	1	49
Library and archives	1	0	0	0	1
Museums and galleries	32	9	6	0	47
Cultural Heritage	2	0	1	0	3
Other	18	3	1	0	21
Total	273	37	35	10	355

If we consider the structure of published texts in VL according to the division of culture, the results are as follows:
- Books and publishing – 66 texts (19% of the published texts). The structure according to the type of text: 45 news (or 68%), 13 short articles (or 20%), 6 articles (or 9%) and 2 interviews (or 3%).

- Visual arts – 13 texts (4% of the published texts). The structure according to the type of text: 7 news (or 54%) and 6 articles (or 46%).
- Music and discography – 103 texts (29% of the published texts). The structure according to the type of text: 82 news (or 79%), 8 short articles (8%), 7 articles (or 7%) and 6 interviews (or 6%).
- Theatre (performing arts) – 52 texts (15% of the published texts). The structure according to the type of text: 46 news (88%), 4 articles (8%), 1 short article (2%) and 1 interview (or 2%).
- Cinematography – 49 texts (14% of the published texts). The structure according to the type of text: 40 news (or 82%), 4 short articles (or 8%), 4 articles (8%) and 1 interview (2%).
- Libraries and archives – 1 text published.
- Museums and galleries – 47 texts (13% of the published texts). The structure according to the type of text: 32 news (or 68%), 9 short articles (or 19%) and 6 articles (or 13%).
- Cultural Heritage – 3 texts (1%).
- Other – a total of 21 texts (6% of the published texts).

Comparison of quantitative indicators of Jutarnji list and Večernji list
Based on the data it is clear that, despite similarities in the newspaper format, differences in numerical indicators of the published texts in the two newspapers do exist. A common characteristic of both papers is the existence of the culture column (where most of the published texts on institutional culture were found). In JL, the culture column consists of an average of 2 pages and in VL the culture column is connected with the stage column and occupies 4 pages (culture itself 2 pages). Such data confirm only one part of the second hypothesis. It is clear that there is no significant difference in the size of the culture column (in JL and VL), but the presumption of an equal amount of published texts was not confirmed. The total number of found texts related to institutional culture was 315 in JL, and 355 in VL. The differences are even greater when it comes to the comparison of texts by type: the amount of news found in VL (273) was almost twice the amount of news found in JL (148). There were more short articles found in JL (100) than in VL (37), and there was also a considerable difference in the number of articles – 60 in JL and 35 in VL. The number of interviews in both papers was almost negligible (10 in JL and 7 in VL).
Regarding the distribution of texts according to the division of culture, there was no major difference in the analyzed newspapers, except in the category of Visual arts (JL published twice as much texts as VL) and the category of Music and discography (VL published 40 texts more than JL).
As a complement to the research results it should be noted that the selected month of research was marked by the long awaited opening of the Museum of

Contemporary Art (cultural event of the year 2009). This cultural event was covered by a significant number of texts. Also, 4 of 6 articles in VL were about the comic book Alan Ford which was being published in their Sunday edition. In other news, short articles and articles no similar patterns appear, and they cover different topics.

Conclusion

It can be concluded that the research (content analysis of JL and VL from 23 November 2009 to 20 December 2009) confirmed the set hypotheses of a low presence of institutional culture in daily newspapers (Hypotheses 1 and 3). Nominally a large number of texts published in Jutarnji list and Večernji list do not reflect the satisfaction with the amount of presence of institutional culture in the analyzed newspapers. Why is that so?

Most texts published in the two analyzed newspapers were only in the form of news and short articles (in JL they constitute 79% and in VL 87% of all the published texts). Their number is large, but, unfortunately, those are only brief information. Despite the fact that every information has its value, there is a lack of serious articles and the question is, how much will be written about culture in the future?

If we look at the structure of the newspapers, it is easy to see that the culture column does exist. However, considering the location, as well as the number of pages it occupies in the newspaper, the question can be asked: "... is the culture in the media truly represented in the right way or does it survive only as a "necessary evil" of the media?"[12]

The traditional order of values in the society, and in the media as well, is certainly different today than it once was. It is the result of a continuous adjustment of the mass media to the modern society and the needs of that society.

"In the past – if we forget sports for a moment– the following order of matters was valid: reliable information on domestic and foreign politics, culture, education and then entertainment. Today, it is like this: entertainment, selective and often unreliable information, more and more pages or minutes of commercial advertising, and – here and there, if there is enough space left – culture and education..."[13]

The overall conclusion of the research is that institutional culture in Jutarnji list and Večernji list is not covered in a satisfactory manner.

---

[12] Mraović, Simo. Intimni odnos kulture i medija. // Kolo. 15 (2005), 2; p. 137.

[13] Lipovčan, Srećko. Je li „2 posto" kulture više ili manje od „pola ure kulture"? // Republika. 62 (2006), 5; p. 71.

References

Croatian Ministry of Culture. Available on: http://www.min-kulture.hr (18 April 2011)

Dragojević, Sanjin; Kanižaj, Igor; Žebec Ivana. Europska Unija u hrvatskim dnevnim novinama : avis – odgoda pregovora. // Politička misao. 43 (2006), 3, p. 133-136.

Kale, Eduard. Uvod u znanost o kulturi. Zagreb : Školska knjige, 1977.

Lipovčan, Srećko. Je li „2 posto kulture" više ili manje od „pola ure kulture"? // Republika. 62 (2006), 5; p. 70-73.

Lipovčan, Srećko. Prinos analizi temeljnih odnosa medija i kulture u hrvatskom društvu. // Društvena istraživanja. 12 (2006), 6; p. 909-925.

Malović, Stjepan. Mediji i društvo. Zagreb : ICEJ i Sveučilišna knjižara, 2007.

Malović, Stjepan. Medijski prijepori. Zagreb : Izvori, 2004.

Malović, Stjepan. Odnos masovnih medija i kulture : globalni trendovi i domaća lutanja. // Kolo, 15 (2005)2; p. 229-251.

McQuail, Denis. McQuail's mass communication theory. London : Sage Publications, 2000.

Moore, Jerry D. Uvod u antropologiju : Teorije i teoretičari kulture. Zagreb: Naklada Jesenski i Turk, 2002.

Mraović, Simo. Intimni odnos kulture i medija. // Kolo. 15 (2005), 2; p. 137-143.

Sadžakov, Slobodan. Marcuseovo shvaćanje kulture. // Filozofska istraživanja. 28 (2003), 6; p. 117-124.

Psychological Aspects of *Cyberspace*

Maja Živko
Department of Information Sciences
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, 10000 Zagreb, Croatia
mazivko@ffzg.hr

Summary

This paper shows that cyberspace has both positive and negative aspects from a psychological point of view. The positive aspects are, temporal flexibility, no limitation of space, the speed of exchanging information, social multiplicity, textual communication, recording ability, entertainment, the empowerment of human potential, equality, and the diminution of the so-called "halo effect". Whereas, the negative aspects are, misinterpretations caused by a lack of sensual integration, the filtering or absorbing of information, frustration caused by technical difficulties, the redefining of identity, "cyberstalking", behavioural disinhibition, and "cyber addiction". Some aspects can be both positive and negative such as, anonymity, distance learning and online psychosocial-globalization.

Key words: cyberspace, self-disclosure, anonymity, asynchronous, synchronous communication, network brainframe, Internet addiction

1. Introduction

In its current usage, the term "cyberspace" is a global network of interdependent information technology infrastructure, telecommunications networks and computer-processed systems. Cyberspace represents "computer-based" channels, "computer networks," and "virtual reality". The term has become a conventional name to describe anything associated with the Internet. This is how the term is understood in this paper. Through this electronic media individuals can communicate, work, create artistic media, read journals, educate themselves, play games, take part in conferences, exchange ideas, etc. In psychology cyberspace is also labelled as the "network of minds". In cyberspace we are able to exchange and share our ideas with others to develop them. In other words, our minds communicate and share ideas with each other. In the "network of reality" there is a dialogical development of meaning. Suler (2007) argues that, "As people view an e-mail, web page or instant message written by an online companion, some people truly feel that their minds are connected to or even blended with the minds of the others".

In the "network of minds" there is an interaction between minds. According to Barak (2008), "cognition in cyberspace is something that happens "between" rather than "inside" subjects and it is a coordinated activity."

Most importantly, the cybernetic brainframe has turned into an "information producer", whilst the individual brainframe has morphed into an "inter-brain-frame".

Therefore, the aim of this paper is to describe both the positive and negative aspects that cyberspace (or the so-called network of minds) has brought into our society from a psychological perspective. It must be stressed that the following aspects discussed below are only one part of a much greater range of issues in this dynamic field.

2. Positive aspects of cyberspace

Clearly the positive aspects of cyberspace are temporal flexibility, no limitation of space, social multiplicity, textual communication, recording ability, the speed of information exchange, entertainment, the unlimited amount of information provided, the empowerment of human potential, equality, and the diminution of the so-called "halo effect".

2.1. Temporal flexibility

In cyberspace time is flexible: communication in cyberspace can be synchronous through chat rooms, in metaworlds where graphical personas or people appear inside a movie set as avatars, or in asynchronous form through the use of e-mails and discussion forums. Asynchronous communication does not require people to interact with each other at the same time. In both asynchronous and synchronous forms of communication (with the exception of video conferencing and Internet-based calls), time is stretched. In an online chat, people have anything from several seconds to a minute or more to reply to another person – this being a significantly longer delay when compared to face-to-face meetings. In e-mails, blogs, and newsgroups, people have hours, days, or even weeks to respond. This gives people time to think about what they want to say and compose their replies in exactly the way they want. This distant communication enables some people to be more expressive, organized, and creative in writing messages.

2.2. No limitation of space

Geographical distance does not exist in cyberspace as computer networks enable people to interact with each other no matter where they are.

2.3. Social multiplicity

Dealing with the relative ease that a person can communicate with hundreds, perhaps even thousands of people. For instance, by posting a message on a blog,

discussion board, or social networking site, which is read by countless numbers of other users, people are connected with other people in an instant.

2.4. Textual communication

Despite the lower sensory quality of text communication, written forms of interaction should not be underestimated as a powerful means of self- expression and interpersonal communication. E-mails, online chats, instant messaging, SMS', and blog's continue to be the most common forms of social interaction. Typing one's thoughts and reading those of another person is a unique way to present a personal identity, perceive the identity of the other online counterpart, and establish a relationship. Some people feel that they can express themselves better in writing.

2.5. Recording ability

A clear and practical advantage of any typed form of communication is that text can always be saved (on disks, tapes), printed or stored as a standard file, which is not possible in reality where people almost always have to rely on their memory.

Thanks to this recording ability**,** people in cyberspace can experience and re-evaluate any part of their communication they wish over and over. They can also use quoted text as feedback to others.

2.6. Speed of exchanging information

In cyberspace there is also the ease and speed with which messages are transmitted across huge distances. People can send and receive messages in a very short time. This is very important, especially in business, where people have to communicate quickly with each other as much as possible.

2.7. Entertainment

The Internet offers entertainment to people in the form of movies, video games, music, online books, magazines, and online chatting. The Internet has therefore become a source of entertainment for millions of people around the world because it offers many cheap and easily accessible forms of entertainment.[1]

2.8. Unlimited amount of information provided

Cyberspace has seen people's ability to better inform themselves improve dramatically, if they should desire. Through a large number of online newspapers, magazines, books and bulletin boards people can find any information they want in unlimited amounts.

---

[1] Entertainment can sometimes have negative aspects, e.g. video games can encourage violence aggression, and fear.

2.9. Empowerment

The Internet also leads to "empowerment", because there are powerful effects from its usage on the way individuals think. We can say empowerment when referring to the full development of human potential, which means freedom from biases, compulsions, hostility, self-doubt, lack of understanding, and an unreflective acceptance of ideologies. The Internet allows individuals to build on their own expertise and ideas. Thus, helping develop creativity and personal identity.

2.10. Equality

The Internet does not discriminate against or marginalize people by gender, class, or ethnicity. It gives people the opportunity to show their own ideas, if they choose too, by creating their own sites on the World Wide Web. Anybody can publish their opinions, needs, interests, and aspirations on the Internet.

2.11. Diminution of the "halo effect"

In cyberspace people cannot judge the personality of another person by his or her physical appearance. This electronic platform distracts us from the effects of any face-to-face interaction, because the perception of one trait (i.e. the appearance or characteristic of a person or object) is not influenced by the perception of another trait (or several traits) of the person or object in question.

3. Negative aspects of cyberspace

Apart from the obvious positive aspects in the online world, there are several negative aspects that need to be mentioned such as a lack of sensual integration, an absorption of information without filtration, any frustration caused by technical difficulties, the redefining of identities, cyberstalking, behavioral disinhibition, feeling that happiness belongs to others, cyber-addiction, etc.

3.1. Lack of sensual integration

Through text-based types of communication the lack of audio-visual sensors such as a change in the tone of our voice, facial expressions, or body language can cause misinterpretations. For instance, tactile stimulation, as well as gustative and olfactive feelings do not exist online.[2] However, cyberspace also allows for various combinations of text, audio, and visual-based communication forms. With so many combinations available, people can hear but not see each other, read texts from each other and see each other but not hear each other, or they can see and hear each other but not bother each other with a written form of communication. In some cases written communication can cause misunder-

---

[2] In cyberspace sensory integration is missing-we see, we hear, but we don't touch, smell or taste.

standings and conflicts.[3] However, in today's multimedia chat environment, misunderstandings and conflicts are lowered because it gives us the opportunity to use an audio-visual form of communication as well. Through the use of various software programs and applications, our voice can be examined more carefully for subtle emotions and meanings that allow people to review their communication between each other.

3.2. Absorbing information without filtering it
Put simply, not all information on the Internet is good, true, valid, or has quality. Individuals can become mere consumers or absorbers of any electronic data flowing through the Internet. These networks lead individuals to expect that they can find a tool that can quickly solve any problem. People impatiently look for technological solutions rather than taking the time to reflect on issues, ponder alternative solutions, and critically analyze problems. Therefore, it is vitally important to filter information and analyze it critically, rather than absorb it passively.

3.3. Frustration caused by technical difficulties
Electronic media is not immune to problems with student-technology interactions, such as any frustrations caused via technical and connection difficulties, and shutdowns. A student's progress during a course, at university for example, can come to a halt due to frustrations with technology. As Wilken (2004) opines "it seems that students become paralyzed until the technical issue is resolved. Students feel abandoned or in some way locked out when they become disconnected from the course due to technical difficulties".

3.4. Redefining identity
Another thing that can occur in cyberspace is a "redefinition of identity." One thing that is particularly troublesome with the formation of identity is the way that young people are being shaped by the Internet. The entertainment and information options that these technologies enable, such as Internet support, enhance, and increasingly define their identities. The Internet can also distort ideas, and manipulate receivers, especially among the youth of today that has a tendency to accept radical ideologies without any critical thinking.

---

[3] Lacking audio-visual effects, an e-mail message, blog, or newsgroup post can be ambiguous. When people read a typed message, there is a strong tendency to project -sometimes unconsciously - their own expectations, wishes, anxieties, and fears into what the other person wrote. Psychoanalytic thinkers call this "transference." Distorting the other person's intended meaning can lead to misunderstandings and conflicts.

3.5. Feeling that happiness belongs to others

Social networking sites (e.g. Facebook, Twitter, etc.) can cause sadness and loneliness. Looking at the "happy" photos of others can lead to some people convincing themselves that other people lead a perfect life and, in turn, they overestimate other people's happiness. Therefore, social networking sites can lead to frustration, sadness and loneliness, as well as the feeling of being a "loser".

3.6. Cyberstalking

Online harassment and cyberstalking are quite common on the Internet. "Cyber-stalkers" can target their victims through several means, including, chat rooms, message boards, discussion forums, and e-mail. Cyberstalking can be occur in a variety of ways, such as sending threatening or obscene e-mails, spamming (in which a stalker sends the victim a multitude of junk e-mail), live online chat harassment or flaming (online verbal abuse), leaving abusive or offensive messages on message boards or guest books, sending electronic viruses or unsolicited e-mail and so on.

3.7. Behavioral disinhibition

It is quite well known that people in cyberspace engage in behavior they would normally deem inappropriate in the real world, due to the fact that they feel more unrestrained. The Internet encourages non-conform, sometimes aggressive, and even unrestrained behavior. The most general cause of problematic behavior is when Internet usage is related to the so-called phenomenon of "behavioral disinhibition." In face-to-face forms of communication, individuals are constrained by the existing social rules that govern interpersonal interaction, immediate negative feedback, and any visible consequences of their inappropriate behavior. Moreover, they are constrained by the social sanctions that would follow after conducting such behavior. In the e-based, and by definition anonymous world, users are often unaware of their identities and personalities and thus can easily ignore the negative consequences of their risky or otherwise unacceptable behavior. This contributes to the expression of anger or aggression and inappropriate self-disclosure.

3.8. Cyber-addiction

Davis (2001) distinguishes between two types of Internet addictions that he divides into specific and general. Specific addictions are an overuse or abuse of any content-specific functions on the Internet (e.g. gambling, stock trading). A general addiction is a multidimensional overuse of the Internet that can result in negative personal and professional problems. A person with a general addiction is drawn to the experience of being online and demonstrates a preference for virtual, rather than face-to-face, interpersonal communication.

For example, the popularity of the social networking site Facebook has led to some serious addiction problems. In psychological literature a term "Facebook addiction" has already been coined. The main issue is that people do not notice the terribly high amount of time and energy that they spend on Facebook. More and more links on web pages invite "sharing" and other types of Facebook-related activities. On the one hand, people use Facebook activities that one could call creative, self-revealing or social, but on the other hand, they could also easily become addicted to it. "Facebook Addiction Disorder" (FAD) is a situation where using Facebook becomes a compulsion and sees people spend all their time logged on to the social networking site whilst ignoring their real social life (friends, family, work.)

4. Psychosocial globalization, e-education and anomymity as positive or negative aspects?

4.1. Online psychosocial globalization

Cyberspace develops web-relationships. Through this new medium individuals can exchange information and find or form groups that share similar interests.

Cyberspace is a "global village or virtual community" (Poole 1997, Wresch 1997). Through social networking sites and virtual worlds (e.g. Second Life, World of Warcraft) people with common interests can share information, work together, tell stories and jokes, discuss politics, help each other, or play games. In cyberspace people also communicate one-to-one and create online groups (e.g. interest and hobby groups, self-help groups, special interest consumer groups). All members of cyberspace have the opportunity to be involved in a society, to which they belong. This is valuable for those who hardly ever communicate with others in the real world, such as seniors whose work and social circles may have become narrowed, shy and timid people, and people who have trouble finding people with similar interests.

Nonetheless, online globalization can also have some negative effects. Cyberspace is exclusively accessible only to those who have the possibility of communicating through e-based media. However, in reality there are still a lot of people with no access to the Internet that are not trained in e-technologies and have no computer skills. On the one hand, there is a global online community, whilst on the other hand, there are those who are disconnected from the Internet world. In this respect, the e-world can also lead to a broadening of existing social inequality.

4.2. Online education

Thurlow and Lengel label a person who learns via computers a "cyber student" (2004:2). The Internet helps cyberstudents gain new skills. Via online books, videoconferences, and discussion forums, individuals can learn about new fields, methods, and acquire new skills. Motivation for e-learning seems to be much stronger than offline learning because of the way of teaching, such as

multimedia platforms, quizzes, games, an easy way to find knowledge sources online, synchronous chatting, discussion forums, asynchronous messages, and e-mail attachments. Individuals can help and support each other through interactions with each other and with their teachers by sharing knowledge and information. In addition, they can also interact with each other through social media platforms such as Facebook, Youtube, virtual social worlds (e.g. Second Life), and virtual game worlds (e.g. World of Warcraft), where people can combine entertainment with education.

In contrast, online education can also have some negative aspects, for example, less liveliness. Classroom discussions can be very stimulating and sometimes much more effective than learning over the Internet. Real-time discussions allow for an instant two-way teacher-cyberstudent interaction and provide the richness of interpersonal interactions so often missing in distance education classes.

4.3. Anonymity

The Internet allows people to conceal their real identity and personal information under false names and identify details which encourage people to express themselves more freely and sincerely than they would in a face-to-face interaction (the so-called "online disinhibition effect). This is because of the fact that they cannot be seen nor heard, which in turn allows people to potentially open up and say things that they normally would not say in the "real world". Thanks to anonymity, self-disclosure and intimacy can be accelerated because people can sometimes disclose more personal, intimate, and sensitive information in the virtual world. In an e-based environment people are more likely to make deeper disclosures about themselves than in reality. Being anonymous in cyberspace is also helpful to those who live in countries where freedom of speech and access to the press are not guaranteed. It creates an open forum to inform others without any danger to themselves.

Anonymity has negative sides too. The physical distance between web participants leads to disinhibited behaviour, because in cyberspace a person can create many identities. Strangers can meet, exchange some right or wrong information and more or less shift their personality so to speak. For instance, a man creates a female identity, or a high school student claims to be an engineer. Internet users are not accountable for what they say, and therefore, positive personal relationships on the Internet are also rare and infrequent.

Conclusion

This article discusses both the positive and negative sides of the Internet world, called cyberspace. Cyberspace allows us to freely choose between synchronous and asynchronous alternative communication, enables us to communicate with thousands of people without the limitations of time and space, present our identity in a written form, record communication, watch movies, play games, listen

to music, read books and magazines, chat, and it empowers our identity to be equal with others. On the other hand, in cyberspace we absorb a lot of information without filtering it and often experience technical difficulties that can cause frustration. We are sometimes manipulated by others on the Internet, or feel that happiness belongs to others. Another dangerous aspect can be e-base stalking by other e-users, called cyberstalking. The anonymity that the virtual world offers can lead us to communicate with dangerous people without being aware of it. Internet users are also globalized in a psychosocial network that leaves those with no access to the Internet disadvantaged and socially crippled. To conclude, cyberspace has become an important part of our everyday life that has both positive and negative aspects that we need to be aware of. Can we ever create a cyberspace world that would function with the lowest possible amount of negative aspects? It is up to all of us to build a new improved future cyberspace where the positive aspects will outweigh the negative ones.

References

Baldwin, E. E. Family empowerment as a focus for home economics education. // *Journal of Vocational Home Economics Education*. 8 (1990)2; 1-12

Baldwin, E. E. Transformative professional practice: Overcoming ambivalence, building community, Kappa Omicron Nu FORUM. 8 (1995),2; 21-37

Basu, Subhajit, J. R. Regulating cyberstalking. //*Journal of Information.* Law & Technology (JILT)2, 2007

Brown, M. M. Philosophical studies of home economics in the United States, Basic ideas by which home economists understand themselves. East Lansing: Michigan State University, College of Human Ecology, 1993.

Brown, M. M., & Baldwin, E. E. The concept of theory in home economics. East Lansing, MI: Kappa Omicron Nu, 1995.

Barak, A. Psychological Aspects of Cyberspace, Cambridge University Press, 2008.

Caroly, S. Wilken. Computer Mediated interaction in a Distance Education. 2004. http://www.kon.org/archives/forum/11-1/wilken.html (2004-08-26)

Cohen, E. Five clues that you are addicted to facebook. 2009. http://edition.cnn.com/2009/HEALTH/04/23/ep.facebook.addict/index.html (2009-04-23)

Cyberspace. 2011. http://en.wikipedia.org/wiki/cyberspace (2011-09-5)

Cyberpsychology, Behavior and Social Networking. 2010. http://www.liebertpub.com/products/product.aspx?pid=10 (2010)

Davis, J. A cognitive-behavioral model of pathological Internet use, 2001.

Gackenbach, J., Psychology and the Internet, Academic Press, 2007.

Green, T. F. The formation of conscience in an age of technology, American Journal of Education, 94 (1984)1; 1-32

Foster, D. Community and identity in the electronic village. In D. Porter (Ed.), Internet culture. New York: Routledge, 1997

Friedman, M. Internet Addiction: A public Health Crisis, 4.22.11

Hadžiselimović, Dž., Plavšić, M., Pregrad, J., Rusijan Ljuština, V. Psihologija mediji etika, Pula: Naklada Slap, 2009.

Haugtvedt, P., Machleit, K., A., Yalch, R., F. Oline consumer psychology. Taylor&Francis e-Library, 2009.

Joinson, A., McKenna, K. The Oxford Handbook of Internet psychology. Great Britain: Oxford University Press, 2007.

Joinson, A., N. Understanding the Psychology of Internet behavior. Palgrave, 2003

De Kerckhova, Lumsden. The Alphabet and the brain. Springer, Verlag, 1988

Journal of Psychosocial Research on cyberspace http://www.cyberpsychology.eu.(2010)

Kimberly, S., Young, Cristiano Nabuco de Abreu. USA: Internet addiction, John Wiley&Sons, 2010.

Libby, C. The Anti-social Network: Is Facebook Making us Sad? 2011. http://www.slate.com (2011-01-26)

Morgaine, C. A., A language of empowerment, Home Economics Forum, 6 (1993)1; 15-20

Nnamdi G. O. Facebook addiction. USA: Ice Cream Melts Publishing, 2009.

Poole, B. J. Education for an information age: Teaching in the computerized classroom. Boston, MA: WCB/McGraw-Hill, 1997

Rehm, M. Empowering the creative spirit: An inquiry into the role of home economics education. //*Journal of Vocational Home Economics Education*. 7 (1982)2; 1-12

Rehm, M. An aesthetic approach to empowerment. Home Economics FORUM, 6 (1993)2; 26-39

Rehm, M. The Internet as a course Practical Problem: Empowerment in the Electronic Global village. 2004. http://www.kon.org/archives/forum/11-1/rehm.html (2004-08-26)

Seel, J. Plugged in, space out, and turned on: Electronic entertainment and moral mindfields. // *Journal of Education*. 179(1997), 3; 17-32

Riva, G., Galimberti, C. The Psychology of Cyberspace. //*New ideas in psychology*. 15 (1997)2; 141-158

Staiger, J. Media reception studies. New York and London: New York University Press, 2005.

Suler, J. The Psychology of Cyberspace Relationships. 2008. http://www.enotalone.com (2008-01-02)

Suler, J. Psychology of Cyberspace - The Basic Psychological Features of Cyberspace. 2005. http://users.rider.edu/~suler/psycyber/basicfeat.html (2005-01-8)

Thurlow, C. C., Lengel, L., Tomic, A. Computer Mediated Communication. USA: Sage Publications, 2004

Walace, P. The Psychology of the Internet. Great Britain: Cambridge University Press, 1999

Wresch, W. A teacher's guide to the information highway. Upper Saddle River, NJ: Prentice-Hall, 1997.

Story Sharing – Model of Collective Collaboration in Online Museum Environment

Željka Miklošević

Department of Information Sciences

Faculty of Humanities and Social Sciences, University of Zagreb

Ivana Lučića 3, Zagreb, Croatia

zmiklose@ffzg.hr

Goran Zlodi

Department of Information Sciences

Faculty of Humanities and Social Sciences, University of Zagreb

Ivana Lučića 3, Zagreb, Croatia

gzlodi@ffzg.hr

Darko Babić

Department of Information Sciences

Faculty of Humanities and Social Sciences, University of Zagreb

Ivana Lučića 3, Zagreb, Croatia

dbabic@ffzg.hr

Summary

The need for a more pronounced role of museums in today's democratic and heterogeneous society has led to a change of the institutional framework which have now come to include voices of all those who have been underrepresented in museums. The new museum needs to pluralize the range of possible meanings by accepting views and interpretation not only of curators but museums users as well. However, museums have been reluctant to adopt new approaches in their work. The paper therefore suggests a more acceptable, on-line model which allows for multivocal interpretations of museum material by connecting the museum data base and web 2.0 applications. The usual repository of actions facilitated by on-line social networks is mostly reserved for web sites outside the direct museum's digital realm. Conversely, the web service presented in this paper supports participative activities on the very museum web site by allowing users to engage with museums' collection. The collection in the form of a digital catalogue can thus serve either as a source or inspiration for individual inter-pretation. Both choosing museum objects from the on-line catalogue and pub-lishing extra-museum material in different media in order to contextualize the

object and create stories builds a two- way communication between curators and actively participating users which is presented to all other types of online users. User-generated content is here seen as a possible influence on the actual institutional policies, namely, collecting, research, exhibiting practices, giving relevance to certain topics etc. By allowing active participation in the creation of meaning, de-authorizing the museum, and at the same time creating a massive depository of heritage information, this model contributes to the paradigmatic shift that has been indicated in the museum world in the last ten years.

Keywords: online database, meaning-production, museum communication, participatory model

Introduction

More than a decade ago, Weil stressed the importance of David Pilbeam's statement that "we see things not as they are but 'as we are'".[1] Indeed, things are meaningful because we assign them meaning, and make these things material culture by investing them with our emotional and intellectual characteristics. Objects created or modified by man "reflect, consciously or unconsciously, directly or indirectly, the beliefs of individuals who made, commissioned, purchased, or used them and by extension the beliefs of the larger society to which they belonged".[2] Material culture has been at the heart of the largest numbers of museums throughout its long history and the shaping of knowledge through material culture, varying in the principles of studies and presentation, has over the last 200 years formed the core of the museum functions as we know them today.[3] Following Foucault's historical nomenclature, Hooper-Greenhill differentiates several modes of knowledge formation in museums, stating that the classical episteme brought a new systematization in the collection according to scientific taxonomies.[4] This positivist age (in the early 17th century) marked an epistemic break with the Renaissance structures of knowing and introduced practices of classification that continued to be followed to the present day. Greenhill's outlook on museum's historical development has a somewhat modified version in Perry's claim that the "age of classification" in museums re-

---

[1] Weil, Stephen E. Rethinking the Museum and other mediations. Washington: London, Smithsonian Institution Press, 1990, 48

[2] Prown, Jules David. An Introduction to Material Culture Theory and Method. // *Winterthur Portfolio*/ 17(1982) 1; 2

[3] Hooper-Greenhill, Eilean. Museums and the Shaping of Knowledge. London: New York, Routledge, 1992.

[4] Ibid

placed curators' individual rationale as late as the 1970s with the introduction of computer based standardization. The late-twenty century ordering and documenting of collections were, according to Perry, shaped by the computer logic rather than just supported by it. "It was perhaps here, therefore, rather than two hundred years before that the culture of the creative 'cabinet' was finally superseded".[5] In other words, computers, or more specifically databases, have become not only the main organizing principle of the modern museum but also a system of thought. This statement is today more valid than ever before since both professional and everyday life of people have been dominated by computer technology, especially in the last decade due to a rapid development of the Internet and web based services. Museums have not only incorporated technology into their work but have also been influenced and modified by it. This paper presents an online generic model adding to a trend towards a changing paradigm of the museum which can be applied as a particular module to the museum's web site and can cover functions that range from the creation of topics to the creation of users' online exhibitions. For the purposes of this paper the authors have chosen to present a specific level of implementation of the model called Story Sharing which enables online users to create their own individual stories and comment those of others.

On a theoretical level, the model includes several major contemporary issues relevant for the discourse on museums, the most significant of which concerns meanings which are construed from objects in museum collections and the parties involved in the construction and distribution of the meanings. The model therefore suggests reconstitution of relationships in the museum, between curators as traditional creators of museum messages (constituting shaped knowledge) and museum users as their receivers. It aims to redefine the (conventional) museum communication process by realizing the potential of the web environment which allows for individual use of the museum collection from the online database and presentation of personal stories. The changing paradigm can be seen in the democratization of the ways in which museum objects are given meanings. In other words, the professional staff of the museum ceases to be the sole producer of information about museum objects. The museum's role here is to enlist visitors as its collaborators who can contribute to the expansion of museum (i.e. curators') knowledge and who can, additionally, develop their own sense of heritage, and create their own links to both an individual and a communal past, but also to the present.

---

[5] Parry, Ross. Recoding the Museum – Digital Heritage and the Technologies of Change. London: New York, Routledge, 2007, 51

Online collections databases – enhancing access and participation

New technology, developing better and innovative possibilities for the distribution of information about the cultural material outside the physical confines of the museum has both followed the change of museum approach to the public and influenced it. It has revolutionized communication between the museum and its audiences in a two-stage development that occurred with the introduction of the World Wide Web – the first bringing accessibility, and the second bringing participation. Computing in museums first changed documentation practices of curators but better information retrieval offered by museum databases slowly ceased to be only curators' privilege. The web environment made possible for the collection to become widely accessible to a great number of people around the world. In an online environment museum material is just a click away. Naturally, visiting the museum building and experiencing the authenticity and originality of artefacts has its advantages. Nevertheless, virtual museums offer alternative possibilities.

After the first stage of publishing online exhibitions, highly curated and linear, museums began showing their collections databases on the web. At first their interfaces resembled the database management system used by museum professionals, but since then, they have developed a highly complex range of options for information retrieval. Searching and browsing is what makes the on line experience of the collection different from the physical museum visit. It is more individual and available to a larger audience and a greater variety of online users.

Ever since they appeared on the web, collections databases of museum objects have been a topic of discussion about whether, and in what form, they could engage properly online users and fulfil their educational and information needs. Donovan sees the first online databases resembling a printed museum catalogue in that they contained object-directed, expert information which was as such, of little interest to the broader public.[6] He therefore proposed that the bare facts of objects be surrounded by layers of interpretation, that is, information stemming from the socio-historical contexts of the objects. In other words, he proposed a content management system that could capture and manage the contents museums create and which they could publish on the Internet. Consequently, the term "access" meant to him only the ability to overcome spatial and temporal constraints to seeing museum objects and reading their labels in a digital form. For that reason he preferred the term "public learning" which could provide better learning opportunities for users through stories presented about objects via "entertaining, prescribed paths that both lead the user lightly by the hand

---

[6] Donovan, Kevin. The Best of Intentions: Public Access, the Web & the Evolution of Museum Automation. // *Museums and the Web 1997*/ Trant, J; Bearman, D. (eds). Toronto: Archives & Museum Informatics

and encourage curiosity, exploration and serendipity".[7] Cameron dealt with similar issues concerning collections databases and the access to the wider content about museum objects.[8] Aware of the emergence of new knowledge paradigms and attempting to redefine documentation practices, she finds that the development of browsing and searching mechanisms could facilitate information retrieval processes. Like Donovan, she argues that collection information should be enriched by both narrative and object-centred histories with which the information can be conceptually expanded. Hyperlinking possibilities offered by the Internet technology, together with the associative systems of meaning making can provide layering and exploration of various contexts of museum objects. A free choice of paths is, in her opinion, inviting for users since they can explore collections in more depth and by self-guided interpretations. By traversing a database and following links between the curator-produced records, users create their own narratives. In combining narratives with object-centred histories Cameron approaches Manovich's view of a database as a cultural form, as a basis for an interface of a new media work. "An interactive narrative (which can be also called "hyper-narrative" in analogy with hypertext) can then be understood as the sum of multiple trajectories through a database".[9] His notion of the narrative belongs to the computerized society. It is the one which replaced the grand narrative of the Enlightenment and became a symbolic form of the modern age, or better yet, computer age, in which it presents the centre of the creative process open to a large number of people. This process of narrative formation is, however, made by selecting already existing information (through links).

Manovich's "hyper-narration", which is actually a form of retelling, can be further elaborated with tools that have been introduced in the second stage of the web revolution that redefined the role of producers in online environments. The development of the so called web 2.0 applications have for a decade now been enabling and encouraging participation from online users who ceased to see themselves solely as consumers of information but started sharing their own experiences and interpretations about certain topics, including those related to museums and their collections. This has been a positive step towards democratization, greater engagement of the public and greater interaction among online museum users through web 2.0 technologies such as blogs, forums, wikis, and

---

[7] Ibid

[8] Cameron, Fiona. The next generation - 'knowledge environments' and digital collections. // *Museums and the Web 2003*/ Trant, J; Bearman, D. (eds). Toronto: Archives & Museum Informatics.; Cameron, Fiona. Object-Orientated Democracies: Contradictions, Challenges and Opportunities. // *Museums and the Web 2008*/ Trant, J; Bearman, D. (eds). Toronto: Archives & Museum Informatics

[9] Manovich, Lev. Database as Symbolic Form. // *Museums in a Digital Age* / Perry, R. (ed.). London: New York: Routledge, 2010 (1999), 69

resulting forms of social networks where museums have created platforms for lively discussion about certain museum events. Openness and user centricity are core components of web 2.0 which have been influencing and encouraging museums to open their strictly controlled collections and change their conventional ways of communications by allowing users to contribute to them.

Collection databases have also been affected by the technological changes. Social tagging or folksonomy[10] has served to encourage user engagement with the collection. "What distinguishes tagging as a form of visitor engagement from other kinds of "interactive" museum programs is that the impetus lies not with the institution but with the individual; the visitor completes the experience. Tagging represents a personal investment in the museum's collection. Visitors add value for the museum, for themselves, and for other visitors by revealing distinct perspectives and connections, and recording them with tags".[11] Adding labels is even a more creative way of engaging users and broadening the interpretative potential of the objects. One such example is the Science Buzz web site of the Minnesota Science Museum.[12] Another kind of user contribution to the meanings of collections has been achieved through wikis. A wiki is a specialized form of Content Management System (CMS) which provides a facility that makes writing to and updating a web site very easy for a group of users and can lead to one or more people building up a corpus of knowledge. The Science Museum Object Wiki[13] was developed in order to engage users with the objects in the museum collection and encourage them to add their personal memories and experiences of using the objects.[14]

Blogs are yet another way of connecting museums and online users, though its organization depends on a temporal ordering of articles and associated comments. Unlike wikis, it allows individual contributions which are shown on the interface as individual, are open to readers' comments to the blog posts. An important element of a blog is the topic. The community that forms online users is highly variable and shaped largely by topic.[15] Museum related blogs can be

---

[10] For example on the website of the Indianapolis Museum of Art http://www.imamuseum.org/art (10 April 2011)

[11] Wyman, B., et al., Steve.museum: An Ongoing Experiment in Social Tagging, Folksonomy, and Museums. // *Museums and the Web 2006*/ Trant, J; Bearman, D. (eds). Toronto: Archives & Museum Informatics.

[12] http://www.sciencebuzz.org/museum/object/2011_01_mummified-hawk/labels (11 April 2011)

[13] http://objectwiki.sciencemuseum.org.uk/wiki/Home (11 April 2011)

[14] Looseley, R., and F. Roberto, Museums & Wikis: Two Case Studies. // *Museums and the Web 2009*/ Trant, J; Bearman, D. (eds). Toronto: Archives & Museum Informatics.

[15] Grabill, J.T. et al. Take Two: A Study of the Co-Creation of Knowledge on Museum 2.0 Sites. // *Museums and the Web 2009*/ Trant, J; Bearman, D. (eds). Toronto: Archives & Museum Informatics

launched by the museum itself[16] or by an online community that feels connected to the museum and its programmes.[17]

Collaborative creation of meanings

Following the developments of online collections databases and interfaces, the Story Sharing online participative model proposed in this paper includes some of the characteristics of the aforementioned examples and shares the lines of thought about the future of museum communication with the broader public. However, it also introduces a new feature that supports the change of knowledge paradigms in museums.

Trying to compare the process of creating hyper-narratives Manovich draws on Mieke Bal's elements of the narrative and states that it is not enough only to create online trajectories. The online user, the creator of the narrative should "control the semantics of the elements and the logic of their connection".[18] Grounded in this statement the Story Sharing model adds the actual creation of information, all performed by an online user. In other words, what this model aims to achieve is to create space(s) on the museum site (i.e. online interface) which will provide online users with a possibility of creating their personal, in-dividual content and share it with the entire online community (Fig. 1). In cre-ating a personal story the user can post on his/her page already existing material by taking it over from the online environment[19], personally created material (texts, photos, videos) and the objects from the online museum collection (im-ages with accompanying text). The created story can be commented and evalu-ated by other online users. Each interface is reserved for only one user, where he/she can add their content and use the museum-created information as well. As such, the model combines the features of blogs and wikis. The key differen-tiating feature is the use of the museum's online collection, in other words, mu-seum objects that can be used and brought into the relationship with individual stories. They become related to unpredictable contexts and find their particular associative place in the "mental maps" of users.

---

[16] Such as the one launched by the Powerhouse Museum in Sydney, Australia http://www.powerhousemuseum.com/collection/blog/

[17] Dulwich on View blog is related to the Dulwich Picture Gallery http://dulwichonview.org.uk/; Liu, A., et al., Dulwich OnView: A Museum Blog Run by the Community for the Community. // *Museums and the Web 2010*/ Trant, J; Bearman, D. (eds). Toronto: Archives & Museum Informatics.

[18] Manovich, Lev. Database as Symbolic Form. // Museums in a Digital Age / Perry, R. (ed.). London: New York: Routledge, 2010 (1999), 70

[19] Provided that the material is not copyrighted or it is published under the GFDL or one of the Creative Commons Licenses

The institutional change suggested by this model is reflected in the way personal stories of users might bring new insights, information, emotional and aesthetic discoveries related to the museum objects and consequently be inserted into the museum database (possible influences on the "upgrading" of the museum databases and on the curators' professional perspectives are shown in Fig. 1 by the dashed arrows)
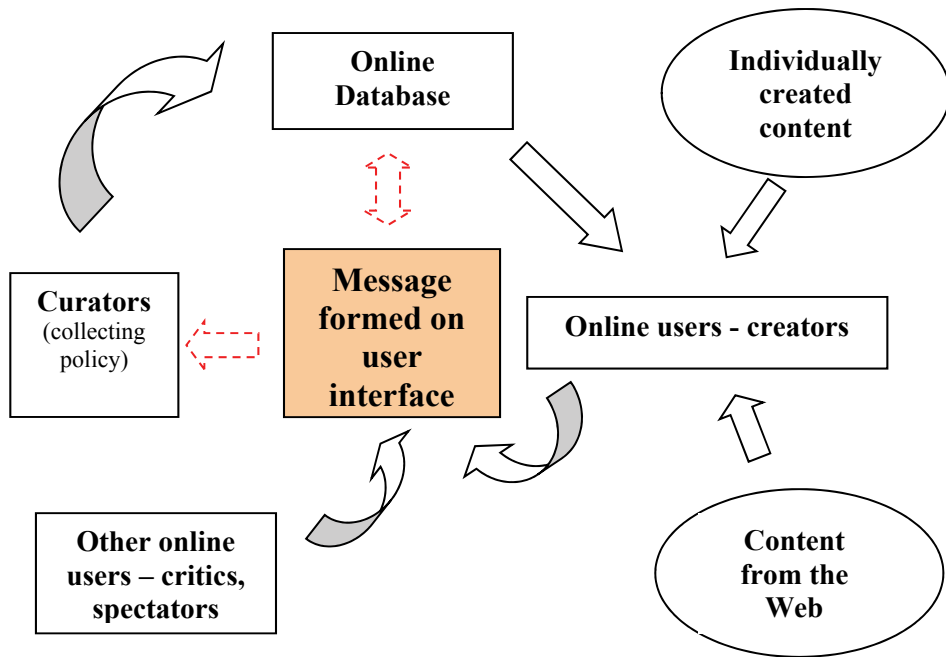
Fig. 1 Creation of personal content related to museum objects and the directions of impact on the institutional "rationale"

In constituting the model it was important to choose the most appropriate form of the desired relationship between museums and their online users. That is why, in addition to spectators, who are seen as the audience in the narrowest sense of the word, the success of the model relies on creators, and to a lesser degree to critics[20]. It is also important to stress that the target users are not specialists since the model does not envisage any sort of information quality or update such as might be evident in wikis. It rather encourages participation that brings into play individual impressions, emotions and/or aesthetic expressions in addition to factual information. It also allows other members of the online

---

[20] Out of all possible categories of online users – creators, critics, collectors, joiners and spectators, http://forrester.typepad.com/groundswell/2007/04/forresters_new_.html

community to post their comments and evaluate stories. The Story Sharing model shows the relevance of the collection not only in terms of presenting curator-produced meanings but of facilitating a feeling of connection and familiarity with the material culture housed in the museum through the production of meanings that stem from the objects' contextualisation within the fabric of people's living memory.

Simulation of the model (html)

The Pool of Topics web page (Fig. 2) represents the central point of the Story Sharing model (within the museum website) where online users (whether museum curators or audiences) can see all the relevant topics relating to or implying the museum's wider sphere of social and cultural engagement.
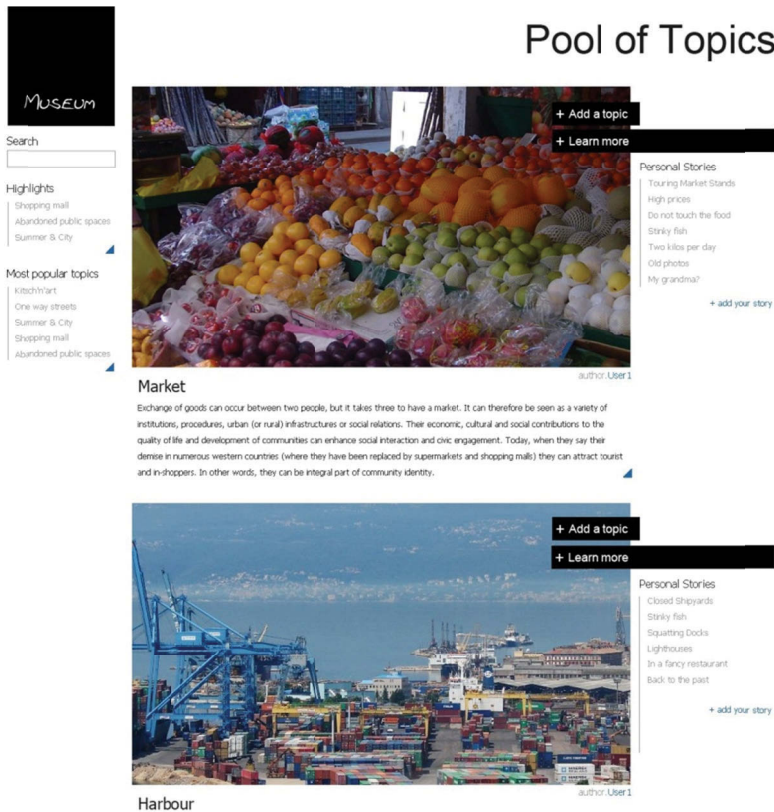
Fig. 2 Pool of Topic web page

In addition to browsing the existing topics, users can suggest and post new ones, but also create their own stories on the topics of their choice. Personal stories (Fig. 3) can be formed by one or more texts and photos and/or videos created by

users and the images of the artefacts from the online museum catalogue. Posted stories are open for comments and assessment by the entire online community. Most important or interesting topics can be selected by the museum curators (which is shown in the left-hand vertical navigation menu, under "highlights"), and a list of the most popular topics crated automatically through filtration of the information about visits and evaluation of certain topics and stories by online users.
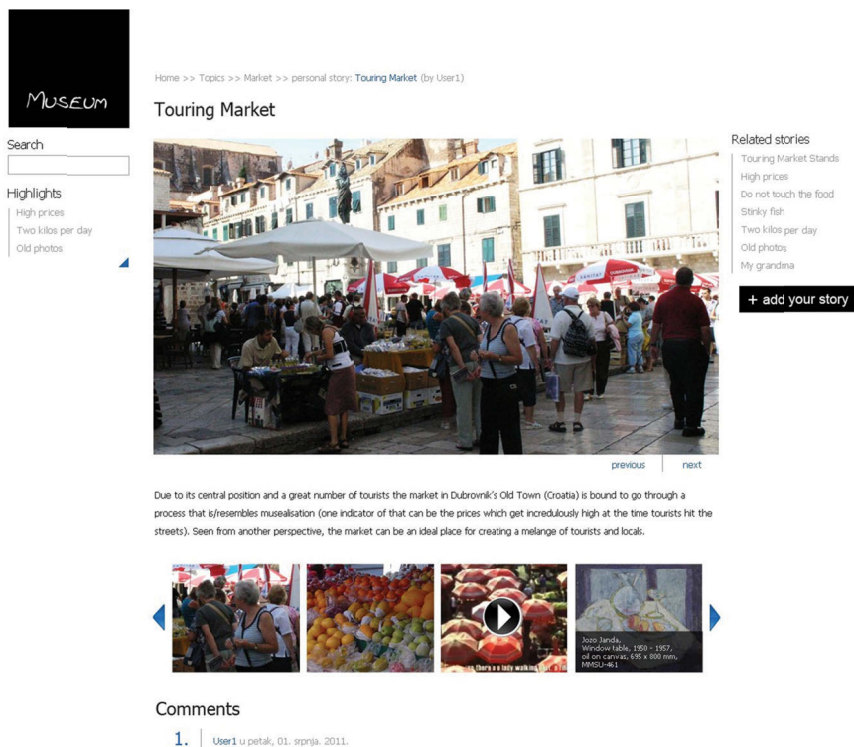
Fig. 3 Personal Stories web page

Concluding remarks

An institutional change encouraged by this model is discernible in several respects. The museum, i.e. curators, assumes a role of a moderator, someone who might choose a topic and moderate discussions and/or different opinions about certain topics. The starting topic is just a set of information on or around which people build their own narratives. The museum also ceases to be the authoritative voice in the dissemination of values but one tone in a cacophony of individual voices forming an online community. As such, the museum can serve as a place, or rather, space for stimulation and empowerment where an online participation is not an end in itself, but a continuous process of the changing of

perspectives of both the institution and its users. The following phase of the project will include qualitative and quantitative methods of evaluating the functioning and success of the model.

References

Cameron, Fiona. The next generation - 'knowledge environments' and digital collections. // *Museums and the Web 2003*/ Trant, J; Bearman, D. (eds). Toronto: Archives & Museum Informatics. http://www.archimuse.com/mw2003/papers/cameron/cameron.html (18 June 2009)

Cameron, Fiona. Object-Orientated Democracies: Contradictions, Challenges and Opportunities. // *Museums and the Web 2008*/ Trant, J; Bearman, D. (eds). Toronto: Archives & Museum Informatics. http://www.archimuse.com/mw2008/papers/cameron/cameron.html (3 July 2010)

Donovan, Kevin. The Best of Intentions: Public Access, the Web & the Evolution of Museum Automation. // *Museums and the Web 1997*/ Trant, J; Bearman, D. (eds). Toronto: Archives & Museum Informatics. http://www.archimuse.com/mw97/speak/donovan.htm#ixzz1PS1D52 H7 (5 July 2010)

Grabill, J.T. et al. Take Two: A Study of the Co-Creation of Knowledge on Museum 2.0 Sites. // *Museums and the Web 2009*/ Trant, J; Bearman, D. (eds). Toronto: Archives & Museum Informatics. http://www.archimuse.com/mw2009/papers/grabill/grabill.html (20 July 2010)

Hooper-Greenhill, Eilean. Museums and the Shaping of Knowledge. London: New York, Routledge, 1992.

Liu, A., et al., Dulwich OnView: A Museum Blog Run by the Community for the Community. // *Museums and the Web 2010*/ Trant, J; Bearman, D. (eds). Toronto: Archives & Museum Informatics. http://www.archimuse.com/mw2010/papers/liu/liu.html (20 February 2011)

Looseley, R., and F. Roberto, Museums & Wikis: Two Case Studies. // *Museums and the Web 2009*/ Trant, J; Bearman, D. (eds). Toronto: Archives & Museum Informatics. http://www.archimuse.com/mw2009/papers/looseley/looseley.html (14 June 2009)

Manovich, Lev. Database as Symbolic Form. // Museums in a Digital Age / Perry, R. (ed.). London: New York: Routledge, 2010 (1999), 64-71

Prown, Jules David. An Introduction to Material Culture Theory and Method. // Winterthur Portfolio/ 17(1982) 1; 1-19

Parry, Ross. Recoding the Museum – Digital Heritage and the Technologies of Change. London: New York, Routledge, 2007

Weil, Stephen E. Rethinking the Museum and other mediations. Washington: London, Smithsonian Institution Press, 1990

Wyman, B., et al., Steve.museum: An Ongoing Experiment in Social Tagging, Folksonomy, and Museums. // *Museums and the Web 2006*/ Trant, J; Bearman, D. (eds). Toronto: Archives & Museum Informatics. http://www.archimuse.com/mw2006/papers/wyman/wyman.html (17 June 2009)

Learning from the Best – What Can Museums in Croatia Learn from the International Museum Practice of Web 2.0

Helena Stublić
Department of Information Sciences
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
hstublic@ffzg.hr

Žarka Vujić
Department of Information Sciences
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
zvujic@ffzg.hr

Summary

Since the first appearance of Web 2.0 in 1999 we have witnessed constant increase in the number of users of its services, especially the society networks and blogs. This trend has eventually encouraged the museums as well to study their possibilities and to start implementing them actively, which is borne out vividly by numerous examples from the world museums. In this paper the most representative ones will be analyzed, i.e. those which make it possible to determine the diverse purpose of using the society networks and blogs – from the promotion of the museums themselves and the cultural heritage they take care of and attracting of real visitors to communicating with virtual users as members of the community within which the museums are active. The latter usage fits into the new mission of the museums and their perception as potential promoters of social changes and active participants in various social processes. The second part of the paper will provide comparative analysis of the examples from Croatia, giving recommendations to the national museum society itself regarding the ways in which actions should be undertaken in order to make maximum use of the mentioned Web 2.0 services.

Key words: social media, museum communication, web 2.0, heritage, virtual community, blog, Facebook

Introduction

Since the appearance of the Internet and the World Wide Web a lot has changed in the global information environment. Today the Internet is a network of different computer networks (network of networks) interconnected in circuits whose dimensions are far beyond imagination. The latest estimates say that in the world today there are more than 1.73 billion Internet users[1]. Along with its development the communication models kept changing as well and they rendered the on-line contents identifiable characteristics. This is particularly evident in the new Web generation, i.e. Web 2.0, which, apart from the standard communication models, promotes also the new, so-called social model where communication occurs among all the participants and in all directions. Based on Web 2.0 the contents like blogs were created and then also the increasingly popular and inevitable social networks. The term blog (blended from Web log) appeared in the 1990s and until today the virtual community of its users has grown into the so-called blogosphere which includes more than 133 million blogs followed according to some estimates by more than 346 million users[2]. Blogs are publications on the web, similar to common websites, characterised by periodical publishing of articles in the reverse-chronological order. They feature all the elements of websites and although the static characteristic of the contents is partly common, through their basic characteristic of being regularly updated and the possibility to communicate with the readers by means of the comments option, blogs already belong partly to the mentioned Web 2.0. The term popular *social networks*, supported by social networking service, covers the combined and well-connected web communities of users that communicate by means of different communication platforms. The best known such web services are Facebook, Twitter, MySpace, YouTube and Flickr. Facebook, the currently biggest social network has more than 600 million active users[2], whereas the number of Twitter users has been estimated at 200 million[3] with a tendency of further growth. Obviously, the potential of Web 2.0 and of social networks has been recognised the same as once the potential of the Web itself. This comes as no surprise considering the main characteristics of Web 2.0 which makes it possible for the user, apart from searching the data and providing information, to actively participate in the design of web contents. Every user of

---

[1] Royal Pingdom. *Internet 2009 in numbers*. http://royal.pingdom.com/2010/01/22/internet-2009-in-numbers/ (21.6.2011)

[2] The Future Buzz. *Social Media, Web 2.0 and Internet Stats*. http://thefuturebuzz.com/2009/01/12/social-media-web-20-internet-numbers-stats/ (21.6.2011)

[2] MSNBC. *Goldman to clients: Facebook has 600 million users*. http://www.msnbc.msn.com/id/40929239/ns/technology_and_science-tech_and_gadgets/ (21.6.2011)

[3] BBC. *Twitter co-founder Jack Dorsey rejoins company.* http://www.bbc.co.uk/news/business-12889048 (21.6.2011)

the service or platform which acts on the Web 2.0 principles can add, delete or change the contents and participate in the discussions on a certain topic. And this is what makes Web 2.0 especially socially attractive, and at the same time powerful as well.

The world of museums and Web 2.0

Over the time the museums have recognised the importance of new technologies and the Internet, but as it very often happens, the adjustment to them and to the resulting new forms of action is slower than in some other social areas. The presence of the museums on the Internet in the form of websites has become almost a must, and there is an evident increase in the development of the presentation of collections and items from the museums and other heritage institutions on-line. The availability of the material in the digital form on the Web is certainly socially desirable, but the design and maintenance are often time and money consuming. Nevertheless, the use of information technologies has brought to significant changes in relation to the communication between the visitors and the museum.

One should keep in mind that the World Wide Web environment covers the following communication models: *one-to-one (i.e. user to user), one-to-many (i.e. user to users – web pages and blogs)* and *many-to-many (i.e. users to users - social networks and wikis)*[5]. The first two communication models are widely present in the museums and they have helped in building and determining the authoritative position of the museums and their respectability and relevance as professional institutions in the field. However, the museums offer also the authenticity, which has been, maybe only seemingly, undermined in the so-called post-museum environment[6] and where communication by means of social networks is based on the *many-to-many* communication model. However, such an approach can result in a stronger on-line interaction and interest of the users for the museums and their collections, as well as for the new interpretative moments and richer dialogue with the audience. However, all this requires a clear attitude and strategy in the museums and similar heritage institutions.

Expressing a clear attitude and design of a clear strategy are in a way hindered by the observed scepticism and discomfort in the interaction with the audience and leaving the communication and interpretation space to the visitors in the way usual for Web 2.0. Reading the recent discussions on this topic, it is difficult to understand the noticed paradox. Regarding their basic function the museums, namely, collect and keep items of the heritage which are, as witnesses of

---

[5] Russo, A; Watkins, J; Kelly, L; Chan, S. *How will social media affect museum communication?*. // 'Proceedings Nordic Digital Excellence in Museums (NODEM)', Oslo, Norway, 2006. http://eprints.qut.edu.au/6067/1/6067_1.pdf (10.5.2011)

[6] Ibid.

123

their time, primarily representatives of a certain time, society or community. This makes the desire to research and present these items in a proper way to the community completely acceptable for the museum experts, and nobody opposes this. However, at the same time the statements can be found such as: "don't feel comfortable in using 'work' time for something that overlaps into a 'social' activity"[7] or "to what extent is the museum willing to promote community knowledge over its own?"[8]. They undoubtedly witness the dilemmas of the museum experts regarding the usage of these new forms of communication with the public. These dilemmas are particularly hard to understand in the publicly owned museums, since they, as public institutions serving the community, should be the ones to recognise the importance of social networks and "social" activities. However, some may say that a high percentage of museums and other heritage institutions in the Western world is of private character and therefore less burdened by performing the social function. Similarly, the mentioned perception of the job and the evaluation of leisure time versus time spent working is problematic, since everything is time consuming. Why would time spent on blogging be less valuable than the time spent on usual curatorial work? These statements show the strong prejudice about what "serious" work represents. The impression is that leisure time and its quality are less important than the time spent on the job. However, the museums are also part of leisure time activity; that is, they are thus qualified within the bureaucratic apparatus. Does this mean that such a qualification makes them less valuable?

Let us mention again, the main components of Web 2.0 are its openness and user orientation[9] which represents new challenges, but also opens up new possibilities for the museums and other similar heritage institutions. The Internet has increased the visibility of museums and their collections and has made them more easily available to the users; however, Web 2.0 has provided preconditions for new communication platforms.

One of them is blog. In fact, blog is currently the most desirable form of Web 2.0 used by the museums. Although first blogs appeared as early as in the 1990s, the first museum blogs were recorded as late as 2002, and we know them as *infoTECMuseo*, *Museum People* and *Modern Art Notes*[10]. Even today the

---

[7] Dicker, E; *The Impact of Blogs and Other Social Media on the Life of a Curator*. J. Trant and D. Bearman (eds). // Museums and the Web 2010: Proceedings. Toronto: Archives & Museum Informatics. Published March 31, 2010. http://www.archimuse.com/mw2010/papers/dicker/dicker.html (25.5.2011)

[8] Russo, A; Watkins, J; Kelly, L; Chan, S. *How will social media affect museum communication?* (10.5.2011)

[9] Ibid.

[10] Spadaccini, J; C. Sebastian. *Radical Trust: The State of the Museum Blogosphere*. J. Trant and D. Bearman (eds.). // Museums and the Web 2007: Proceedings, Toronto: Archives & Museum

museum blogs are rare on the Web, witnessed by the fact that according to 2007 estimate[11] there are only two hundred of them. This in itself shows the fact that museums unwillingly and with difficulty opt for risky and unpredictable project with unknown effect and uncertain outcome[12].

On the other hand, experts emphasise that the advantages of a blog are great, that they are easy to start and use, inexpensive and least damaging to the current information policies of the museums. The museums can use blog, as well as social networks, to go beyond the frames of their community and to become part of the global events. In this they have to satisfy two conditions – they have to have a constant connection to the Internet and be capable of maintaining their web contents in the English language if they do not belong to the English speaking area. It is a well-known fact that the English language is the *lingua franca* of the new age and that it is considered the official language of the Internet world. Due to this fact the bi- and multilingual websites and blogs have become everyday part of the Web.

In the first place, blog can serve to connect with the current visitors and users, by exchanging ideas and information and by stimulating various discussions. In this case the communication is facilitated by the very fact that the medium of the blog, as well as of the social network in general, is based on informal approach and it is optional for the participants. The users find it easier to provide a comment if this may be done anonymously or using a nickname, but even in this case the author of the blog still keeps all the rights of editing and deleting the comments. Furthermore, blog is an excellent means of promoting museums and their activities, and may be used for expansion and attracting new users and potentially future visitors. The fact needs to be kept in mind here that a great part of the social network users belongs to the younger population which is sometimes part of the museum non-visitors. They require a specific approach, responding well to the Web 2.0 rules – informality, openness and innovativeness – which is why these forms of communication can only be of help to the museums in attracting this problematic part of the potential public. The presence of museums on the social networks and the existence of museum blogs provide a new public image of museums as institutions which keep pace with new technologies, close to the young, who understand their language and are not afraid to get involved in the projects with immeasurable results, all this in order to be closer to the members of their community.

Regarding responsibility for the museum blog, by analysing the existing examples we have noticed that in some museums this is the job of one person only. It

Informatics, published March 1, 2007. http://www.archimuse.com/mw2007/papers/spadaccini/spadaccini.html (25.5.2011)

[11] Ibid.

[12] Ibid.

is usually the blog of the institution itself, having the character of an official channel for the communication with the users. Not so often there are museums with several blogs edited by several persons, usually curators interested in specific topics, as well as the description of collections and jobs they perform. Some museums use the approach in which every project, regardless of whether it is an exhibition, workshop or any other specific project has its own special blog. The most striking examples are the blogs that record the creation of temporary exhibitions. It happens often also that the institution has one blog that represents a museum, with several persons, usually curators, being responsible for its management, i.e. writing and publishing articles on the blog. In this way the continuity of publishing articles is ensured, as well as the diversity of their contents and a larger span of topics creating the feeling of unity and togetherness within the institution.

Letting the curators manage the blog does not necessarily mean that this is additional work for them, but rather it is considered part of their professional work. This means that they are expected to independently use the necessary services and to make the entire process, from the idea and writing of the article to designing and publishing it on the blog, on their own, since by understanding the social networks and by active participation in the entire process, each participant can best approach the task, feel the "pulse" of the network and act in compliance with the users' requirements and their own preferences and discretion. In this way the curators can also experience the entire project of blog management as their own project, and this may in turn motivate them to get involved with greater dedication. It should also be mentioned that in cases when several departments of a museum manage the blog (e.g. curators, IT department or Web department) the entire process is slowed down and is subject to unpredictable and unnecessary complications.

When starting a blog its purpose and basic topic need to be well considered. In principle, one may speak about two main types of museum blogs. One, that primarily addresses its community, and the other type of blog is of a more professional nature and addresses the museum professionals. One classification mentions also blog as curator's portfolio where the curators present certain aspects of their work and interesting facts bringing closer in this way the curator's job to the public, at the same time rendering the museum a human note.

Museum blogs are often of limited duration, which is in fact rather far from being characteristic of the blog form. The reason for this lies in the fact that they have been designed exclusively in the function of one project, usually, as already mentioned, in the function of temporary exhibitions which are excellently promoted and popularised by the specific insight into the process of their development and very interesting *'behind the scenes'* approach on the blog. The users play here a very important role since on the blog they can provide interesting ideas or suggestions while the exhibition is in the phase of being created and

126

thus they can participate in its front-aid (content level) and formative evaluation (display elements).

The attendance rate of the blog, including its success, are maybe most affected by the frequency of publishing new contents, regardless of whether this refers to articles, photo materials or videos, because these keep the readers constantly interested. The diversity of blog topics is also an important factor which ensures greater attendance rate and more permanent readers, but the main idea and the blog topic should be always kept in mind. Publishing of two articles (i.e. posts) per week proved to be optimal, whereas rare updating and displaying of new contents leads usually to the reduction in the number of active readers. One should also keep in mind which is the potential and target audience so as to form the topics and contents in compliance with their needs and objectives. The blog will be the more successful if it manages to reach its audience and if it establishes the dialogue with it. One of the parameters of a successful blog is also the number of comments realised per article. Referring to a blog as a marketing aid should be made with caution since this is not its primary task. If blog, as well as any other social network, is used exclusively as means of promotion, they very soon become uninteresting to the users and thus fail in their purpose. Smartly designed, interesting and well visited blog is the best marketing an institution can wish for.

Apart from the blog, various **social networks** have proven as excellent. The advantages of social networks such as Facebook include easier access to the users' opinions, simpler presentation of the museum activities and a kind of low-budget marketing. For social networks many equal rules and pieces of advice are valid as those mentioned for the blogs. The difference lies in the fact that on social network the basic idea is not regular publishing of articles, but rather provision of short pieces of information, sharing of interesting links or initiating discussions on the forum. Social networks serve for strong connecting among people who share common interests. One of the frequent objections to social networks is their unpredictability in operation, changes they bring and the number of users. The museums cannot predict in which ways the social networks and the information published on their services are going to be used, what number of people participate in a social network and which is its scope, which results in the impossibility of making any long-term planning of activities or of the duration of participating in such an environment. Because of many dilemmas, the lack of research in the area of social networks and their overall unpredictability there is need to a kind of *radical trust*. Radical trust is a term which refers to the trust that the institutions must have towards the rules of conduct and the operation of social networks when they decide to use them. In this sense Darlene Fichter says: "*We don't have a million customers/users/patrons … we have a million participants and co-creators. Radical trust is about trusting the*

community….. As an online community we come up with safeguards or mechanisms that help keep open contribution and participation working"[13].

Examples of using Web 2.0 in museum practice

As already mentioned, statistically looking in relation to the overall number of blogs on the Web, there are few of the museum ones. However, it may be said that there is quite a number of them which are of high quality and interesting. This leads to the conclusion that although the museums need a long time and a lot of motivation to get involved at all in such a project, those that eventually opt for it, take this job very seriously and invest a lot of efforts in order to keep the blog at a high quality and respectable.

In compliance with the earlier typology of the communication model on the Web, several examples will be given that stand out by the contents or organization. The first one is *Eye Level*, the Smithsonian American Art Museum blog which in the description, almost in the manner of the museum mission explains the representative mission of the blog which is quoted here in full: "*The name Eye Level imparts a sense of clarity to which the blog aspires. The name refers to the physical experience of viewing art, but it also plays on the many roles and perspectives that make a museum a reality—roles that will come into focus here. (...) Using the museum's collection as a touchstone, the conversation at Eye Level will be dedicated to American art and the ways in which the nation's art reflects its history and culture. The discussion will extend beyond the walls of the Smithsonian American Art Museum's collection to include other collections, exhibitions, and events. Eye Level will also document the extraordinary collaboration between curators, conservators, handlers, historians, enthusiasts, critics, exhibition and new media designers, and of course bloggers that has motivated the past and present of American art history*"[14]. All this shows that the blog belongs to the group of museum blogs which serve to present the museum and its work, based on the desire to communicate and discuss the American art among the blog readers as well. The blog is managed by several curators who publish new articles relatively often and regularly. Each article provides key words, i.e. tags or labels, to facilitate the search, and it is equally possible to divide it into nine different social networks. However, let us be sincere and say that such a blog can be established in America precisely within the Smithsonian, the only museum group at the federal level that is particularly keen on an excellent public image.

In the well-known London Victoria & Albert Museum with a long tradition they opted for a different approach. They organize, namely, special blogs for every

---

[13] Russo, A; Watkins, J; Kelly, L; Chan, S. *How will social media affect museum communication?* (10.5.2011)

[14] Eye Level. http://eyelevel.si.edu/ (4.6.2011)

major exhibition or project. Thus on the blog *Creating the Cult of Beauty* all the details can be read about the respective exhibition, from the organisation and the entire process of design to the very opening and finally closing that is planned in July of this year. In his latest article on the blog the curator of the exhibition says: "*It is hard to believe that* The Cult of Beauty *has now been open to the public for two months. I still enjoy going down to the galleries and hearing the reactions as visitors see the exhibition for the first time, as they take their first glimpse into Rossetti's bedroom, step into the projected Peacock Room or come almost face to face with* Eros. *I thought I would use this, my final blog post, to suggest some other Aesthetic buildings and events you might like to add to your itinerary. To echo the words of many online retailers, 'If you enjoyed The Cult of Beauty you might also like to try....*"[15]. The blog therefore acts in the typical 'behind the scenes' manner with numerous comments and positive reactions.

An example of blog which is not institutional, but can be called a museum blog is *Museum 2.0*[16] managed by Nina Simon, a well-known museum exhibition designer and author of the book 'The Participatory Museum'. Her blog addresses primarily the museum community, and deals with a wide spectrum of museum topics. Nina Simon publishes articles minimally once a week, has a large number of readers, responses and comments and is an excellent example of a blog that acts as a platform for information and ideas exchange, offering at the same time also a place for discussion.

Unfortunately, the museums in Croatia have not yet shown courage to start their blogs and therefore there are no national examples to present.

When speaking of social networks, Facebook leads regarding the number of users and therefore, any major museum in the world has its open so-called page within Facebook which, like *profile* for individual users, serves the organisations, institutions, etc. as a form of participation. There is also the option of *group* which can be founded by one or several Facebook users with the aim of connecting with the people sharing the same interests. The popularity of a group is seen in the number of group followers who receive regular reports (i.e. news feed) on their *wall* about new information sent by the group. In Croatia the situation on social networks is much better than the blogger one. The museums in Croatia tend to accept this type of communication more. The preference to social networks over the blogs can be explained by the fact that forming of groups or profiles on Facebook requires minimal time and is completely free. Sending information, invitations or links is very simple and efficient. Several dozens of museums in Croatia out of two hundred of the existing ones can be found on

---

[15] Creating the Cult of Beauty. http://www.vam.ac.uk/things-to-do/blogs/creating-cult-beauty (4.6.2011)

[16] Museum 2.0. http://museumtwo.blogspot.com/ (4.6.2011)

Facebook, but the fact that the entire process of accepting and using social networks in Croatia is still in its experimental phase is confirmed by the fact that the majority has been registered as individual users, i.e. *profile*, and not as *page* which is the option intended precisely for the institutions and organisations.

Conclusion

The aim of this paper has been to illustrate the current condition of using Web 2.0 and its tools in the museum environment. In spite of many drawbacks of the Internet communication and Web 2.0 and their uncertain development, everything mentioned leads to the conclusion that the advantages and benefits of blogs and social networks in the operation of museums are multiple.

Blog or social network can present a big gain both for the museum as an institution and for every individual museum professional. Regular and high-quality communication with the users can help the curator to remain focused, be aware of the world around and some other and new perceptions, and even different attitudes. It provides help in thinking and looking *outside the box*, which ensures that the results of the work remain interesting and relevant to the community and the society in general.

Also, the usage of these tools allows communication and connection among museum professionals in the entire world. This is precisely what makes in fact the museums, not only their collections, but rather their employees as well. Providing them with the possibility of exchanging information, experiences and ideas, and the insight into the best practice, it will be easiest for them to learn, develop and act creatively. Such employees will make the museum an institution recognised and loud in the community, the one which is not only an active participant in the social processes, but rather also an initiator of new and important initiatives and projects.

However, by including the museums into the Web 2.0 community the winners will also be the users. Their communication with the museum experts will be facilitated and they will have faster, easier and less emotionally or psychically burdened access to certain information, and maybe even more important, they will have opportunity of expressing their opinions, criticism, suggestions or desires. Also, a group on the social network such as Facebook can connect the users among themselves who, by participating in this group, expectedly, share the same interests.

And finally the most important issue – the informality of communication provided by blogs and social networks allows bringing the users closer to the museums and recognising the museums as friendly and easy-to-access places for everyone. The museums, on the other hand, are provided with better fulfilment of the known mission of being oriented to and serving the users, the community and the society in general.

References

BBC. Twitter co-founder Jack Dorsey rejoins company. http://www.bbc.co.uk/news/business-12889048 (21.6.2011)

Creating the Cult of Beauty. http://www.vam.ac.uk/things-to-do/blogs/creating-cult-beauty (4.6.2011)

Dicker, E. The Impact of Blogs and Other Social Media on the Life of a Curator. J. Trant and D. Bearman (Eds.). // Museums and the Web 2010: Proceedings. Toronto: Archives & Museum Informatics. Published March 31, 2010.
http://www.archimuse.com/mw2010/papers/dicker/dicker.html (25.5. 2011)

Dawson, B; et al. Social Presence: New Value For Museums And Networked Audiences. J. Trant and D. Bearman (Eds.). // Museums and the Web 2008: Proceedings, Toronto: Archives & Museum Informatics. Published March 31, 2008.
http://www.archimuse.com/mw2008/papers/dawson/dawson.html (25.5. 2011)

Eye Level. http://eyelevel.si.edu/ (4.6.2011)

Gates, J. Case Study: New World Blogging within a Traditional Museum Setting. J. Trant and D. Bearman (eds.). // Museums and the Web 2007: Proceedings, Toronto: Archives & Museum Informatics, published March 1, 2007
http://www.archimuse.com/mw2007/papers/gates/gates.html (25.5.2011)

MSNBC. Goldman to clients: Facebook has 600 million users.
http://www.msnbc.msn.com/id/40929239/ns/technology_and_science-tech_and_gadgets/ (21.6.2011)

Museum 2.0. http://museumtwo.blogspot.com/ (4.6.2011)

Royal Pingdom. Internet 2009 in numbers. http://royal.pingdom.com/2010/01/22/internet-2009-in-numbers/ (21.6.2011)

Russo, A; Watkins, J; Kelly, L; Chan, S. How will social media affect museum communication?. // 'Proceedings Nordic Digital Excellence in Museums (NODEM)', Oslo, Norway, 2006. http://eprints.qut.edu.au/6067/1/6067_1.pdf (10.5.2011)

Spadaccini, J; Chan, S. Radical Trust: The State of the Museum Blogosphere, J. Trant and D. Bearman (eds.). // Museums and the Web 2007: Proceedings, Toronto: Archives & Museum Informatics, published March 1, 2007
http://www.archimuse.com/mw2007/papers/spadaccini/spadaccini.html (25.5. 2011)

The Future Buzz. Social Media, Web 2.0 and Internet Stats.
http://thefuturebuzz.com/2009/01/12/social-media-web-20-internet-numbers-stats/ (21.6.2011)

Digitising Contemporary Art (DCA)

Jerica Ziherl
Museum of Modern and Contemporary Art
Dolac 1/II, 51000 Rijeka
jerica.ziherl@mmsu.hr

Nataša Šuković
Museum of Modern and Contemporary Art
Dolac 1/II, 51000 Rijeka
natasa.sukovic@mmsu.hr

Summary

The DCA project aims to create high-quality digital reproductions to assure the long-term preservation as well as online access to such reproductions and their data. By July 2013, Rijeka's Museum of Modern and Contemporary Art, as one of the DCA partners, will digitise 730 artworks selected from their own collection. The metadata on the digitised objects will be mapped, enriched, contextualised, and aggregated for ingestion into Europeana. The partner institution can then make this digital content available through their own websites and other channels as well. This will facilitate the public online access to (parts of) contemporary art collections of partner institutions and foster augmented user experience. The DCA project intends to enhance the online visibility of contemporary art as an essential expression and an invaluable building block of European culture.

Key words: digitization, contemporary, art, museums, metadata, preservation

Introduction

The Museum of Modern and Contemporary Art (MMSU) was one of the first in Croatia which had have constructed its own digital collection of museum items supplementing it by online segments. Since 2005 a database emerged containing almost 7000 entries in respective collections divided on diverse visual art segments that through the online catalogue provided a global accessibility to the Museum's Collection in its virtual form. Despite MMSU personnel DIY principle, at that time it was an innovative project that introduced the institution to a contemporary communication forms and global dissemination of information on status and value of visual culture in Rijeka and Croatia. Aware of the global cultural policy of the „Digital Renaissance", launched on the large scale by the European Commission within projects like *Europeana* or *The Information and Communication Technologies Policy Support Programme* (ICT- PSP) and re-

flected on the national level by the *Croatian Cultural Heritage* (CCH) project, MMSU continuously aims to foster greater cooperation between its role as the regional actor who is looking ways to strengthen the links between the various institutions involved in improvements and creation of up-to-date digital content, especially of contemporary artworks - a kind of cultural heritage still largely missing from *Europeana* and from CCH project on the national level.

Therefore, it was of great importance to join Belgium *PACKED vzw* initiative born out of wish to valorize the expertise that emerged out of the best practices in the field of conservation and presentation of Digital Art over the past two decades. The prevailing thesis that new media art is defined by its conceptual content and physical effects to a greater extent than it is by its original material- ity provided the backdrop of the debate and recent efforts have resulted in the articulation of a new theoretical framework and new methods for digitalization.[1] At the same time, the reports on the results of an extensive research project about Installation Art led to a new terminology and offered a much better ap- preciation of, nowadays, one of the dominion artistic genre carried out by an international group of custodians active in the conservation of contemporary art.[2] Installations can be endless, often to the despair of the custodian of the work. How can one preserve and document the installation? What relation ex- ists between the components and the space, and what is the spectator's part in the work? Questions of this kind are examined in connection with a number of case studies. Furthermore, main European projects on Digital and Installation Art put an emphasis on ethical questions, e. g. IPR clearance which is still in its infancy considering visual arts in Croatia. So on *PACKED vzw* initiative MMSU entered a newborn project -*Digitising Contemporary Art* (DCA). And whenever new project starts, new rules are drafted, new questions arise. In fur- ther text we are not only explaining this statement, but also, by following the same structure, helping the reader to locate where he/she may find the path in the project and contribute to the practical and theoretical discourse on this issue.

About the DCA

Digitising Contemporary Art, launched on January 1 2011, is a 30-month digiti- zation project for contemporary art, precisely the art made after 1945. It is co- financed by ICT – PSP programme and intends to enhance the online visibility of contemporary art as an essential expression and an invaluable part of Euro- pean culture. Thereby DCA wants to stimulate the interest of the general public by introducing a stronger presence of contemporary art as a reflection on the art of today to the *Europeana* portal. The aim of the project is not only to fill the gap in *Europeana's* content supply but to develop the best practice in digitiza-

---

[1] See the Variable Media Network, IMAP, DOCAM, GAMA and mediaartbase.de among others.

[2] See the project *Inside Installations* (2004-2007).

tion, preservation and conservation of such complex fields such as contemporary art. By comprising paintings, drawings, photographs, sculptures, installations, objects, videos, films etc., DCA will create a digital corpus of 26.921 high-quality reproductions of artworks and 1.857 of contextual documents which will become accessible and retrievable through *Europeana*, providing metadata and thumbnails that will serve as direct links to large-sized reproductions of each item on partner institutions portals. DCA corpus includes masterpieces from key artists of most European countries and will contain texts and images, as well as video and sound material – both still underrepresented on the portal.[3] The artworks and contextual documents mainly belong to institutions whose collections are less known or still unknown and that need support for their digitization. The 21 collections come from 12 European countries: 17 of which are from countries that are behind in making their heritage accessible through the European cultural heritage portal and 2 countries which are lagging in their effort to make their cultural heritage accessible. The main issues to be addressed within the project are the choices of specifications for digitization and metadata, so that they may be inter-operational, and finding the appropriate aggregation solution for each institution. DCA digitization action will also contribute to the preservation of the artworks. The digital images produced in the context of DCA will become part of the digital collections of each contributing institution. They will care for their long-term sustainability, as they do for their other data and images. DCA itself will provide guidelines and assistance on how to preserve digital files and keep them accessible over a long period of time.
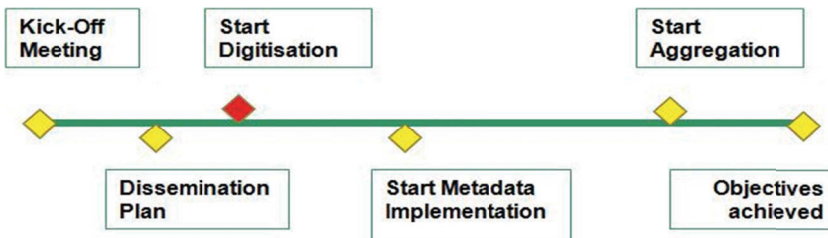
Picture 1. DCA project work-flow

About the DCA Consortium
Competitiveness and innovation framework of the DCA project was flawlessly prepared by PACKED vzw. As a Project management who overall legal, contractual, ethical, financial and administrative management of the project will create and support the conditions necessary for a successful and effective col-

---

[3] Some of its best-known artists are Marina Abramović, Orla Barry, Christian Boltanski, Marie José Burki, Gusztáv Hámos, IRWIN, Sanja Iveković, Bjorn Melhus, Carsten Nicolai, Dan Perjovschi, Fiona Tan, Blast Theory, Luc Tuymans, Steina Vasulka, Franz West.

laboration and performance within the 24 partners from 12 countries; some were key partners in previous European projects. List of consortium is divided as follows:

Table 1. Content providers – museums and art institutions involved in the project (divided by countries)

Austria	Ars Electronica, http://www.aec.at
Belgium	argos - centre for art and media (Brussels), http://www.argosarts.be MAC's - Museum of Contemporary Art of the French Community of Belgium (Grand-Hornu), http://www.mac-s.be Mu.ZEE – Collection of the province of West Flanders and the City of Ostend (Ostend), http://www.muzee.be Royal Museum of Fine Arts of Belgium (Brussels), http://www.fine-arts-museum.be
Croatia	MMSU - Museum of Modern and Contemporary Art Rijeka (Rijeka), http://www.mmsu.hr
Germany	EMAF – European Media Art Festival (Osnabrück), http://www.emaf.de HfG – Staatliche Hochschule für Gestaltung Karlsruhe (Karlsruhe), http://www.hfg-karlsruhe.de Transmediale (Berlin), http://www.transmediale.de
Greece	Frissiras Museum (Athens), http://www.frissirasmuseum.com MMCA - Macedonian Museum of Contemporary Art (Thessaloniki), http://www.mmca.org.gr National Gallery-Alexandros Soutzos Museum (Athens), http://www.nationalgallery.gr
Island	National Gallery of Iceland (Reykjavik), http://www.listasafn.is RAM - Reykjavík Art Museum (Reykjavik), http://www.artmuseum.is
Latvia	Latvian Centre for Contemporary Art (Riga), http://www.lcca.lv
The Netherlands	Museum Boijmans Van Beuningen (Rotterdam), http://www.boijmans.nl NIMk - Netherlands Institute for Media Art (Amsterdam), http://www.nimk.nl
Poland	WRO Art Center (Wroclaw), http://www.wrocenter.pl
Portugal	Serralves (Porto), http://www.serralves.pt
Slovenia	MG – Moderna Galerija (Ljubljana), http://www.mg-lj.si
Spain	Antoni Tapiès Foundation (Barcelona), http://www.fundaciotapies.org

Technical partners

Greece	NTUA - National Technical University of Athens (Atena, Grčka), http://www.ntua.gr
	Ubitech – Ubiquitous Intelligent Technical Solutions (Atena, Grčka), http://www.ubitech.eu
Belgium	Multimedia Lab Ghent University - IBBT Interdisciplinary Institute for Broadband Technology (Ghent, Belgija), http://multimedialab.elis.ugent.be

The selection of artworks

During the selection period each artwork from Museum's Collections was reviewed based on several criteria. They included considerations of presentational aspects of digitalized items within the project in relation to the extension of user population, as well as possibilities of creating new values by connecting artworks on the contextual level. Primarily, decision was to include the works that have representative status within Croatian contemporary art scene, showing initial points and logical developments of certain artistic tendencies, parallel to development of European and world art. Subsequently, the selection included works made by internationally renowned artists, recognized in critical and professional circles, as well as by regional artists representing innovative and experimental approaches. Finally, the criteria was to include works of cultural, historical and art-historical importance reflecting the artistic preoccupations, as well as cultural, social and political issues the other half of the 20th and early 21st century, characterized by a high aesthetic value and high artistic methodology. However, additional criteria soon appeared, focused on one of the basic functions of the institution - the preservation of artworks that have not yet been digitized and whose digitization is a priority in museological terms. After careful review, Collection curators have chosen overall 100 works from the Collection of Drawings, 100 works from the Collection of Prints, 150 works from the Collection of Photography, 10 works from the Collection of Posters, 290 from the Collections of painting, 20 from the Collection of Sculptures and 60 works from the Installations and New Media Collection. Thus, a total of 730 works that represent overachievements in their specific medium in relation to the specific context in which they appeared, providing a uniquely comprehensive survey of artistic tendencies in contemporary art.

The problems of IPR clearance

Since the artworks that we aim to digitise within this project have been created after 1945, almost all of them are still protected by copyright. However, since *Europeana* operates by a „clean hands" policy, all the participating collecting institutions had to ensure that the copyrights on all content will be cleared and that *Europeana* will be able to display metadata and previews of the artworks,

as well as links to high-quality reproductions within their original context (on the website of collecting institution). It was very difficult to give an estimate of possible outcomes of the process of IPR clearance considering our contracts are rather diverse in terms of restrictions regarding creation of copies, distribution and presentation of materials. There are also few cases where there are no contracts or documentation of entry of the Museum's items, only the Purchase Rewards documentation. In these cases works were, especially after the major international exhibitions, left at the Museum and any attempt to contact the artist regarding regulation of the status of their work failed to be completed so, after a certain time, they were simply merged in the Museum's holdings. Moreover, often the distribution rights intended for the promotion of the works include only the right for displaying the work as well as the right for publishing the reproductions of work in classical printed material. It should be taken into account that these contracts are made at a time when digitization of works for on-line presentation was not a common museum practice. Finally, some contracts for the works made in co-authorship were signed only by one author without special authorization which would guarantee the agreement of others.

For these reasons, prior to digitization, we started the demanding process of updating contacts of the authors, co-authors and their right holders, as well as updating the Museum's database with new information about the IPR holders after the author's death. We are currently in the process of making new versions of the contracts or annexes to the existing contracts that will involve more elaborate conditions for presentation and access to digitized content.[4] In accordance to DCA consortium recommendations, we are planning to implement specialized contracts for photographers who will perform the digitization of the material, which will include the terms of use of digitized content and quality assurance of digitization in accordance with predefined guidelines and parameters.

The preparatory phase - MMSU digitization plan

The first six months of the project was envisaged for project preparation activities that included reviewing the proposed materials and determination of their characteristics based on a few basic metadata such as: inventory number, title, year, author/artist, type (e.g. sculpture, painting), the material/carrier (e.g. oil/canvas, video/DVD), size/format, the current location (e.g. storage in a museum), where will it be digitised (e.g. Museum photo studio or outdoor location), transportation requirements, the condition of work (i.e. whether there is need for restoration of the works), IPR owner (the artist/artist's family), whether

---

[4] Project leader PACKED has, together with the University of Patras (Greece), developed an online step-by-by guide for museums and other culture heritage organisations with regards to IPR issues within the project *ATHENA – Access to Cultural Heritage Networks across Europe*.

the rights have already been cleared for digitization/on-line presentation. For audiovisual works some additional questions were required related to the producer of the work, status of source material (e.g. master copies, presentation copies), rights we hold on source material (e.g. non-profit distribution/presentation within the Museum's activities, preservation, documentation), duration and other technical parameters (e.g. mono/stereo, the number of channels, video/film, black/white or color). These metadata improved the creation of tree-year digitization plan by providing the necessary filtration system for extensive material and influenced setting of objectives, resource planning and selection of technologies and specifications for individual operations.

Table 2. Content table

ITEMS TO BE DIGITISED	
Item no.	MMSU-2467 (1-160)
Title	Greenhouse
Year	2003
Artist (s) / Creator	Brajnović, Tomislav
Producer (in case of AV/sound work)	The Artist (in case of Film and Video)
Type	Installation
Material & Carrier	Greenhouse - glass and metal construction; various used items (ready-mades and found objects), film, video, sound, light, fog
Status of source material (in case of AV/sound work)	Screening copy (Audio), Masters (film and video)
Rights you hold on source material (in case of AV/sound work)	Rights for public presentation
Sizes / format	220 x 223,7 x 264 cm
Other technical parameters (in case of AV/sound work)	Film: (MMSU-2467-148), 8mm, colour (also stored/copied on Beta digital record and on DVD) Video: (MMSU-2467-147), VHS, colour, (also stored/copied on Beta digital record and on DVD) Audio: (MMSU-2467-146), CD (copy from the LP), stereo
Duration (in case of AV/sound work)	Film: t=2,14 min/sec Video: t=32,32 min/sec Audio record: t=52 min
Current location	External housing facility
Where will it be digitised	MMSU, Rijeka
Transportation required	Yes – 500 m and Artist's assistance
Condition of the work	Good
IPR owner	The artist
Rights already cleared for digitisation	No, but will be cleared during the DCA project
Rights already cleared for online publication	No, but will be cleared during the DCA project

Picture 1. Tomislav Brajnović, *Greenhouse*, ready-made greenhouse and used items, the standard 8 mm film, video, sound, light, 220 x 223,7 x 264 cm, 2003, MMSU-2467 (1-160)

Setting of objectives remained present during the whole decision-making process, relating to each individual work while respecting the overall structure of digitalization plan. Two objectives were constantly kept in mind while developing the digitization plan:

a) conservation goal:
substitution, e.g. creation of digital master copies that will serve as a substitution for analog video due to fragility of the original (such as VHS tapes) or obsolete carriers (e.g. VHS-recorder),
documentation for conservation purposes, e.g. creation of high-quality digital master copies of works that are in common use to minimize frequent handling which diminishes their condition.

b) access/presentational goal:
research, e.g. creation of high resolution digital master copies that will allow large increases and provide a detailed study of brush strokes in paintings,
on-line presentation, e.g. presentation of collections to on-line users (usually in lower resolution).
After analysis and breakdown of each proposed work to the components and their characteristics in accordance with predetermined objectives, the works have been re-grouped into categories within the digitization plan separated by time segments. The plan was supposed to include clearly specified objectives and products of the project, plan of activities, responsibilities, resources and evaluation procedures.

Do it once, do it right!

The saying that often runs in the guidelines of this project is *Scan one for every purpose*, or variations *Do it once, do it right!* It is applied as a metaphor for all the types of digitization, from scanning and shooting to video digitization. Therefore, it is proposed that delivery includes high-quality digital copies (master copies) from which the user copies are made that try to anticipate every potential use of digital reproduction (e.g. printing, web). As "best practice", it is recommended that digital files created by photographing the items should come in 5 different formats:

- .dng, for digital negative: this open format comes a standard for the .raw format, that always depends on the camera type (there are in fact dozens of ".raw" formats); storing a .raw is useless if you don't have the proper camera to operate it; the .dng avoids this problem;
- Uncompressed baseline IBM TIFF v6.0 CMYK, min. 300 dpi in A2 format, without interpolation; this comes as the "best" size, but it is not mandatory; the CMYK file will be used for printing (see below: output), and compare the result with the original;
- Uncompressed baseline IBM TIFF v6.0 RGB, same resolution: the RGB file is the best for viewing on a screen (see below: output); yet, to ensure perfect colors display, it should be used on a calibrated screen (not mandatory).

The following two formats are better suited for access purposes:

- .jpg RGB 300 dpi in A5 format: for intern use on the Intranet, for instance;
- .jpg RGB 72 dpi in A6 format: low resolution, to be published on the Internet.

The main obstacle for delivery of these technical parameters makes a request for A2 format with a resolution of minimum 300 dpi without interpolation that can be delivered only by cameras that use multi-shot technique, recording several consecutive photos that are later integrated into one. However, more important than the camera itself and its chromatic quality is, of course, primarily a photographer with a good feeling for characteristics of the original (e.g. color, two-dimensionality of images, spaciousness of installations), and the whole process that follows after the shooting. Nevertheless, if we want to be sure that delivery corresponds with the original and that we haven't vainly spent the budget for low quality reproductions, we must engage in a color correction of printed material. Otherwise, the most common scenario is to simply store digital copies in the storage until at one point they are needed for catalog printing and after a few months or years you get unpleasantly surprised by the fact that the digital reproductions have totally different colors than the original. Photographers then claim that the problem is in printing, publishers shift the blame to photographers and artists then blame us. It is therefore necessary to check the quality of digital

reproduction and its CMYK version printed on calibrated printer according to ISO 12647v2 standard and its comparison with the original in a standardized viewing condition in order to avoid all effects of metamerism (the effect of light on colour pigments). Color corrections require an investment in the training of museum staff, purchase of necessary software, printers, lighting and is there-fore, time and money consuming.

The three-year digitization plan for MMSU

The three-year digitization plan for MMSU consists of four segments divided into four basic categories that relate to IPR clearances, metadata, digitization and the online implementation. Each of these activities within the segments are elaborated individually and include time frames for each item: creation of file names, updating, coding and implementation of metadata, the preparatory work (restoration, cleaning of VHS tapes, transport organization, participation of artists in set up of their works), creation of conservational masters and derived user copies (technical parameters and time frame in which they will be real-ized), post-production ("cropping" photos, leveling posters), quality control of deliverables (control files for viruses, testing formats, color corrections), storage of digital content, marking carriers (e.g. hard disk, Digital Betacam) and the implementation in Museum's website and *Europeana*.

Picture 2. The three-year digitization plan for MMSU

Conclusion

Regardless to different levels of use of digital content enabled by the DCA pro-ject, different user groups often operate in a similar way - everybody wants an easy and quick access to trustworthy, high-quality digital reproductions of con-temporary artworks – and if they have no commercial intentions, then they will also expect to get it free of charge. Anyone will be able to view the newly cre-ated digital content on the Web. In *Europeana* (and other portals) they will be pointed to the original context of the found items (e.g. the museum's website) by a link, in order to consult enhanced visual data and additional information on the actual work. Such links will enrich the *Europeana* experience, increasing the visibility of the contributing museum's website and also encouraging part-ners to develop (apart from DCA) new web applications, e.g. tools that will al-

low the user to create their own virtual exhibitions or collections. This will result in a more interesting experience of the institutions'contemporary art collections for the public.

References

Coelho, Ramon; Rodrigo, Evert; Wijers, Gaby. The sustainability of video art: Preservation of Dutch video art collections. Amsterdam: The Foundation for the Conservation of Modern Art, 2003

Frasco, Lizzie. The Contingency of Conservation: Changing Methodology and Theoretical Issues in Conserving Ephermal Contemporary Artworks with Special Reference to Installation Art. Philadelphia: University of Pennsylvania, 2009

Kimbell, Lucy. New Media Art: Practice and context in the UK 1994-2004. London: Arts Council England/Cornerhouse, 2004

Manovich, Lev. The Language of New Media. Cambridge: MIT Press, 2001

Paul, Christiane. The Myth of Immateriality: Presenting and Preserving New Media. // *Media Art Histories* (ed. Oliver Grau). Cambridge: MIT Press, 2007

Rosen, Margit; Schön, Christian. From Time Based Arts to Database. Frankfurt: Revolver, 2003

Scholte, Tatja; Wharton, Glenn. Inside Installations: Theory and Practice in the Care of Complex Artworks. Amsterdam: University Press, 2011

Zilić, Diana. Many hand make light work. // *Formal Engaged: from the MMSU collection* (ed. Branko Franceschi). Rijeka: Muzej moderne i suvremene umjetnosti, 2008, 80-94

E-SERVICES, E-GOVERNMENT AND BUSINESS APPLICATIONS

E-Government and e-Participation:
City Web Sites – the Case of Croatia

Domagoj Bebić
Faculty of Political Science, University of Zagreb
Lepušićeva 6, Zagreb, Croatia
domagoj@edemokracija.hr

Milica Vučković
InMed Institute
Šibenska 1, Zagreb, Croatia
milica@edemokracija.hr

Summary

Relying on the Kang and Gearhart study about the content-specific relationship between citizens' use of city web sites and civic engagement, this study examines the relationship between specific content categories on city Web sites and civic engagement in Croatia. Building upon Kang and Gearhart's theoretical arguments about motivational communication theories, web site functions, and civic engagement in a new technology context (e.g., Coleman, Lieber, Mendelson, 2008; Jeffres and Lin, 2006; Norris, 2003), this study examined the role of the city web sites in e-participation by surveying citizen users of city web sites. A web based volunteer survey was conducted in 33 Croatian cities, with more than 600 citizens participating. Furthermore, building on Norris's concept of democratic divide we have content analyzed the web sites of cities that participated in the online survey. The results of content analysis of the city web sites indicate that there is democratic divide (Norris, 2001) in Croatian local government. Secondly, the comparison of the findings obtained by the public opinion web based survey with the results of the content analysis study supports Scott's (2006) findings about city web sites' potential for citizens' civic engagement. Thirdly, the results support Kang and Gearhart's findings which exhibit a clear indication that city web sites can actively promote civic engagement among citizens. In other words, this study demonstrates that citizens are willing to engage online if the government provides them with an adequate platform.

Key words: e-participation, civic engagement, e-government, city web sites, direct democracy, democratic divide

Introduction

Due to the rapid rise of information and communication technologies (ICTs), governments are more and more willing to provide the public with online delivery of services and programs. The way citizens are using technology has changed. Governments have had to adopt new modes of behavior as well; they had to do what the business and civil sector had already been doing for a long time: bring services and programs online. Moreover, they were expected to use the potential of the internet to reinforce the public sphere and engage citizens. Today it is hard to imagine a government that did not bring at least some of the services and programs online, which does not have at least a web site. Baring in mind that a web site is seen as a local portal for community services (Ho, 2002), a space for information services and interconnectivity with citizens (Musso, Weare, & Hale, 2000), a platform for citizens' engagement, it is argued that via these web sites, citizens often have a public commitment to, knowledge of, and interest in issues, places, and social networks that enable public decisions and actions (Bowles & Gintis, 2000; Jeffres & Lin, 2006). Furthermore, experts find that cities and towns are usually moving faster in e-government service delivery than their national counterparts. The explanation is that local governments are closer to their citizens in terms of services needed.

Kang and Gearhart building upon theoretical arguments of motivational communication theories, web site functions, and civic engagement in a new technology context (e.g., Coleman, Lieber, Mendelson, & Kurpius, 2008; Jeffres & Lin, 2006; Norris, 2003) in their study assessed the role of city web sites engendering civic engagement by surveying citizen users of city web sites. Their findings suggest that surveillance; practical services and direct democracy features function as important conditions for citizens' e-participation (Kang and Gearhart, 2010). They came to the conclusion that city web sites can improve e-participation. Furthermore, relying on the Schumpeterian framework, Norris finds that parliamentary web sites should serve two primary functions that reflect the liberal notion of parliament in representative democracy: the top down channel that stands for providing the public with information, and a bottom up channel of communication from the public to the elected members.

Building on these findings we want to examine is there a democratic divide in Croatia. Secondly we look at the relationship between specific content categories on city web sites and civic engagement in Croatia. Moreover, content analyzing the city web sites, we want to see what features does a specific city web site have and which of those features are most often used. While Kang and Gearhart were looking only at the citizens' dimension, in our study we will make a comparison of the findings obtained by the public opinion web based survey and the results of the web sites content analysis.

In the first chapter we bring the arguments of authors who see the internet as a medium for broader political participation. Moreover, we support the idea that on-line technologies can serve as platform for reconnecting citizens and the

government. Secondly we discuss the role of city web sites as mass communicators. Hereby we rely on Jeffres and Lin's (2006) theory that city web sites should include mass communication functions: surveillance, coordination of activities, socialization, and entertainment to fulfill the goals of civic engagement. After this theoretical part, in the second chapter we bring our research design, the last chapter deals with results and discussion. The conclusion summarizes the core findings and gives ideas for further research.

Theoretical part

Stephen Coleman noted that the democratic deficit is facing many governments and that two key questions arise from that: "How to make the political process more participatory; and how can public engagement in policies that affect everyday life become more deliberative?" (Coleman 2003). Hence, Coleman believes that the new ICTs could contribute to a renewed faith in the government bodies through the creation of a more transparent, interactive government engaged in wide dialogue with an interactive citizenry. Substantial literature speculates that the Internet could strengthen civic engagement and political activism, especially for many groups currently marginalized from mainstream politics (Norris, 2000). Norris summarizes: "…the Internet offers to reconnect people to the political process by helping people become more informed citizens, by helping representatives become more responsive to citizens, and by engaging more people in public policy debates (2000: 3)." In this context Norris talks about the democratic divide defining it as divisions between those able to use the internet for enhancing their political participation and influence (Norris, 2001). Norris argues that parliamentary web sites should function in a way to provide comprehensive information about their activities and interactive communications to encourage public feedback if they are to fulfill their democratic function (Norris, 2000).

Relying on the Schumpeterian framework, Norris finds that parliamentary web sites can provide the public with necessary information about legislative procedures and activities, allowing greater transparency and promoting the accountability of elected members to their constituents (2000: 6). Furthermore, parliamentary web sites can also provide additional 'bottom up' channels of communication for citizens seeking to contact elected representatives, via email, online discussion groups, straw polls or other feedback mechanisms like comments pages, which advocates of direct democracy regard as important mechanisms of public deliberation and participation (ibid).

Direct democracy theory posits that democracy works best when citizens are directly involved in policy debates, decisions, and actions (Barber, 1984; Coleman & Gotze, 2003). Authors who believe the internet is a new remedy for disengaged citizens, believe that web sites can serve as platform for engaging citizens in policy making (Hague & Loader, 1999). Norris (2003) and Scott (2006) suggest that possible indicators of direct democracy in city Web sites are all in-

dicators from representative theory and pluralist theory, online issue discussion forums, e-consultations, voluntary service opportunities, and virtual meetings. However, direct democracy also requires government policies and programs that reflect the collective knowledge and commitment of citizens (Norris, 2003). Hence, it is argued that successful civic engagement largely depends on effective public communication between a city and the citizens via various communicative means ranging from meetings to city web sites. Furthermore, Hale et al. (1999), stress that the web sites can be a relevant channel for citizens' civic involvement including citizen discussion on and participation in city improvement activities. Secondly, a city web site is an arena for e-democracy (Hague & Loader, 1999). In other words, city web sites can act as a civic mobilizer, encouraging citizens' involvement, and having an impact on city government improvement and city development (Musso et al., 2000). City web sites provide citizens with information about public and political affairs and elections, which can amplify citizens' interest in political behaviors (Kang and Geahart, 2010).

As we can see, authors find that city web sites have the potential to provide citizens with a platform for civic engagement. However little was done to investigate the connection between specified city web site content and civic engagement. Yet, Kang and Gearhart, working in motivational communication theories framework, tried to answer this question. They bring an example of the uses and gratification theory, which explains why people use media overall (McQuail, 1985). Shah (1998) found that television viewing resulted in civic engagement depending on the programs the audience watched. Even though overall television viewing reduced civic engagement, viewing of social drama facilitated civic engagement. However, viewing science fiction did not predict civic engagement. Relying on these findings, Kang and Gearhart found in their study specific indicators contributing to civic engagement: information about city services, coordination of activities and direct democracy features rather than surveillance functions are likely to induce civic involvement, secondly, interactive services including access to city meeting agendas, administrative services, online forums, and prompt feedback lead to citizens' active participation in political development (2010: 458). To sum up, the authors find that "city web sites offering quality of life, cultural aspects, and citizen input opportunities have potential to encourage citizens' participation in city development (ibid)."

City web sites and civic engagement are discussed also in the role of e-government as a mass communicator. Jeffres and Lin argue that city web sites should include mass communication functions: surveillance, coordination of activities, socialization, and entertainment to fulfill the goals of civic engagement, wherein they can play the roles of surveillance by providing information about city service programs, budget, or annual reports, by offering connection with city government they provide coordination function, providing news about the city, arts and cultures, recreation and entertainment, information about civic organizations and tourism, neighborhood, health, and links to local media they

can fulfill socialization and entertainment functions. Performing these functions, city web sites play many of the same functions as mass media web sites, what makes them mass communicators. Jeffres and Lin (2006) address that a city web site performing mass communication functions, where citizens can find virtual networks for information and participation can provide citizens with the arena for civic engagement.

Research design

As we have discussed, little is known about the relationships between citizens' city web site use and civic engagement. Therefore, the current study posits the following two research questions.

RQ1: Is there democratic divide in Croatian local government?

RQ2: Which content categories of city web sites are more often used by citizens?

To answer our first question we conducted a content analysis of available city web sites. We analyzed 31 city web sites for 26 items. Since there are 110 cities in Croatia in total; the number of 31 cities participating in the survey was relevant. We included in the analysis all the cities members of the City Association that also had publicized a link on our web survey. Among the examined 26 items, in the literature, six of them are defined as features for e-democracy. According to Kang and Gearhart, these are: discussion forums, online voting, application forms about volunteering, online consultation, online opportunities for citizen involvement in policy making, online opportunities for citizens' virtual meetings, feedback link for contacting mayor online and city council reps. Additionally, in this group of variables we have included and link on official profiles on social networks, because it is argued that social networks can increase e-engagement, especially among younger population. An inter coder reliability test was conducted on 10 web sites and reached high score at 0.99

Secondly we conducted a voluntary web based survey. Participants in our study were citizen users of city web sites in Croatia. The data was collected by conducting a web-based survey publicized on more than 30 city web sites which are members of the City Association. Web masters of the city web sites were asked to post the survey link to their web sites, and were sent two follow-up e-mails to remind them of the survey link. The survey was conducted from June 1 to July 1, 2011. A total of 33 cities out of 105 in the City Association posted the link, a 34% participation rate. Citizen respondents were asked to provide their city name in the survey, and the results reported citizens from a total of 33 different cities participated in the survey. The number of citizen participation varied from 1 at minimum, to 226 at maximum per city. Once participants agreed to participate in the survey, they could continue the survey. The online survey questionnaire consisted of questions representing three democratic theories (representative, pluralist, and direct democracy) and the mass communication functions (surveillance, coordination, entertainment, and socialization) in a Web

technology context, civic engagement, media use, demographics, and political orientation (Kang and Gearhart, 2010). The survey was programmed to allow only one completion per respondent. The total number of participants during the 1 month period was 681. After eliminating respondent dropouts, final sample consisted of 541 participants. Bagozzi, Dholakia, and Pearo (2007) contend that the nature and extent of response bias for this kind of web survey are unknown. Nonetheless, as the number of citizen participants is large and they represent many cities, this convenience sample was deemed relevant for answering the research questions posed. The result support Johnson et al.'s findings that web users for public affairs or government information are likely to be males, older than 30 and highly educated, politically left oriented and interested in politics (Johnson et al., 2008).

Considering the services and information available in city web sites, this study employed measures of civic involvement and political behaviors to measure civic engagement (Kim, 2007; Moy et al., 2005). Civic involvement items included a total of five items, and political behaviors measured responses to five questions asking their participation in political activities in their community. Citizens' use of city web sites was gleaned from previous studies about city web site content. Relying on Kang and Gearhart's research, representative features such as a connection to the mayor, pluralist features such as links to local civic organizations, direct democracy items such as online forums, and mass communication features such as surveillance, entertainment information, and other overlapped features with democratic theory were used to measure city web site content categories for civic engagement. Citizens were asked, how often, in the past 12 months, they used a variety of content in their city web site representing democratic theory and mass communication functions.

Control Variables. For control purposes, the survey measured demographics, political orientation, and media use variables. Demographic variables included gender, age, education and income. Citizen respondents were also asked about their general interest in politics and political ideology. Media use was evaluated as the degree of each medium use (television, radio, newspapers, magazines, and the Web).

Results and discussion

Our first RQ asked if there is a democratic divide in Croatian local government. Content analyzing web sites of 33 Croatian cities we found that there is democratic divide in Croatia. Accordingly, among 26 examined features, we coded 7 of them for direct democracy; online consultations, virtual meetings, discussion forums, feedback links, social networks, online voting, and online opportunities for civic involvement in policy making. Findings reveal that 13 cities have only one of these categories, and in most cases that category is „feedback link to contact the Mayor or public officials", while rest of the cities have two or three categories, excluding the city of Rijeka, which has five categories and the city

152

of Split which does not have any. As Graph 1 demonstrates, the direct democracy function is not fulfilled and the potential of the internet as platform for engaging citizens online in policy making remains unfilled. However, the analysis reveals that the city of Rijeka is a definite leader in Croatia when it comes to engaging citizens online. If we compare these findings with the number of participants in our survey, we are free to suggest that a city web site can serve as platform for engaging citizens online. Moreover, the findings from the web based survey, which will be discussed in the next paragraph, demonstrate that citizens of the city of Rijeka in more than 50 percent of examined cases use direct democracy categories on the city web site, as well as other web features. On the other hand it is surprising to find that Croatia's second biggest city Split does not have any of the categories that serve for direct democracy.

Graph 1, demonstrates that other web functions, surveillance and practical services are well fulfilled. This brings us to conclude that local governments in Croatia still see city web sites as platform for the "top down approach". Particularly, as the analysis reveals, they provide citizens with information about social services, they bring reports from each city council and they have uploaded budgets (surveillance). Secondly, on more than 60 percent examined cities one will find tourist information, news stories, information about transport, city history, information about zones and city inspections etc. (practical services), which means that local governments still see web sites in the function of a mass communicator.

Graph 1: City web functions
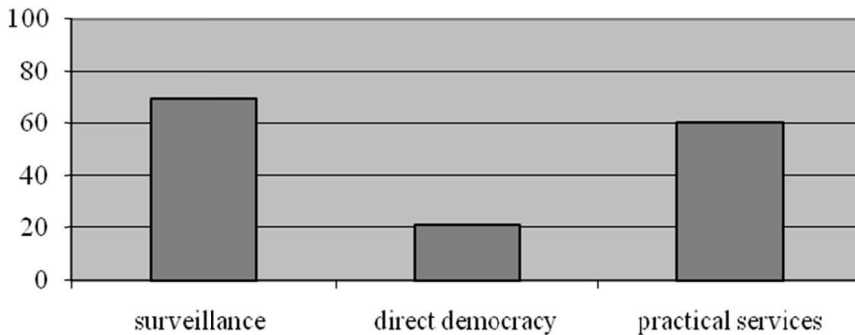
Second RQ was supposed to answer which categories of city web sites are more often used, from which point we could see which categories stand for civic engagement. Analyzing answers, the results reveal that citizens are willing to participate online and use available services. 76 percent of participants answered that they visit the city web site often or very often. For instance, participants used online voting once or more in 33 percent of cases, and content analysis showed that 39 percent of examined cities had online voting. We have looked

for example which cities have forums about city issues and compared that with the use of that item. The results demonstrate that participants in our survey use city forums in a great measure, for instance, the citizens of Rijeka used the forum once or more in 64 percent of cases, citizens of Velika Gorica in 88 percent, of Zadar in 75 percent, Vukovar 71 etc. We have similar findings when looking at online voting. The content analysis demonstrates that each examined city has news stories about the city, and 86 percent of participants in the survey answered they had visited this category on the web site once or more. All the city web sites in the survey have the category of city history which has been visited once or more by 84 percent of citizens.

An example that supports these findings is the city of Rijeka. The city of Rijeka had all 26 items that we were looking at. Accordingly, more than 220 citizens of Rijeka participated in this survey. In-depth analysis demonstrates that among these citizens more than 50 percent of them used the items on the web site which are related to direct democracy.

Conclusion

Building on the notion that city web sites should serve as platform for a top down approach, as well as for a bottom up approach, and that different features of a public web site promote civic engagement, by content analyzing city web sites and looking at the results from a web based public opinion survey, this paper developed several assumptions. First, conducting content analysis of city web sites, we argued that there is democratic divide in Croatian local government. We believed that local governments in Croatia still see a web site as a panel on which they will put all information and documents available for the public, but will not provide citizens with platforms for engaging in policy making. In other words, as Norris (2000) formulates it, they maintain only top down approach, from the government to the public, but they miss the approach from the public to government, they miss to give the citizens an opportunity to see what they think and how they feel about certain issues. The results support our assumption. Direct democracy features are utilized only up to 21 percent (Graph 1), while other functions of city web sites are reached at higher level, above 60 percent, for surveillance and practical services.

Secondly, we have conducted a voluntary web based survey. We asked the citizens about their civic involvement, political behavior, and, for our study the most important, questions related to the civic use of city web sites. We asked them how often they used certain categories on city web sites. We wanted to see if there are any categories for which we could say that they are positively associated with civic involvement. Coleman and colleagues for instance found that the use of tax information and connection with city officials in the web site was associated positively with an attitude toward civic engagement (2008). However, our study did not reveal any significant findings at this point. The fre-

quency test demonstrates that we cannot isolate any category for which we could say that it improves citizens' engagement.

Finally, we argued that citizens are willing to engage in the decision making process if the platform for that is available. Comparison of the results from the content analysis and the web based survey supports this notion, which is best demonstrated with the example of the city of Rijeka. The city of Rijeka fulfilled direct democracy and mass communication functions of a city web site in a great measure. And not just that, citizens recognized that. From the results of the web based survey we can see that they use those items often. Furthermore, it is relevant to mention that more than 70 percent among 224 participants from Rijeka feel they are a part of the city.

To sum up, new media and internet have changed the ways in which individuals interact and communicate with each other. In return, we got citizens who are expecting more from governments in the way of technological solutions and platforms for engaging them.

References

Bagozzi, R. P., Dholakia, U. M., & Pearo, L. R. (2007). Antecedents and consequences of online social interactions. Media Psychology, 9(1), 77–114.

Barber, B. R. (1984). Strong democracy. Berkeley: University of California Press.

Bowles, S., & Gintis, H. (2000). Social capital and community governance. Retrieved April 13, 2008, from http://www-unix.oit.umass.edu/_bowles.

Cap Gemini Ernst & Young. (2003). Online availability of public services: How is Europe progressing? Retrieved April 20, 2008, from http://europa.eu.int/information_society/europe/ 2005/doc/highlights/whats_new/capgemini4.pdf.

Coleman, S., & Gotze, J. (2003). Bowling together: Online public engagement in policy deliberation. London: Hansard Society.

Coleman, R., Lieber, P., Mendelson, A. L., & Kurpius, D. D. (2008). Public life and the internet:If you build a better website, will citizen become engaged? New Media & Society, 10(2), 179–201.

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook ''Friends'': Social capital and college students' use of online social network sites. Journal of Computer-Mediated Communication, 12(4), 1143–1168.

Hague, B., & Loader, B. (1999). Digital democracy: Discussion and decision making in the Internet age. London: Routledge.

Hale, M., Musso, J., & Weare, C. (1999). Developing digital democracy: Evidence from California municipal web pages. In B. Hague and B. Loader (Eds.), Digital democracy: Discourse and decision making in the information age (pp. 96–115), New York: Routledge.

Ho, A. T-K. (2002). Reinventing local governments and the e-government initiative.   Public Administration Review, 62(4), 434–444.

Jeffres, L. W., & Lin, C. A. (2006). Metropolitan websites as urban communication. Journal of Computer-Mediated Communication, 11(4), 957–980.

Johnson, T. J., Kaye, B. K., Bichard, S. L., & Wong, W. J. (2008). Every blog has its day: Politically-interested Internet users' perceptions of blog credibility. Journal of ComputerMediated-Communication, 13(1), 100–122.

Kang, S. and Gearhart, S. (2010). E-Government and Civic Engagement: How is Citizens' Use of City Web Sites Related with Civic Involvement and Political Behaviors?, Journal of Broadcasting & Electronic Media, 54: 3, 443–462

Kim, S. (2007). Media use, social capital, and civic participation in South Korea. Journalism and Mass Communication Quarterly, 84(3), 477–494.

McQuail, D. (1985). Gratifications research and media theory: Four models or one. In K. K. E. Rosengren, P. Palmgren, & L. Wenner (Eds.), Media gratification research: Current perspectives (pp. 149–167). Beverly Hills, CA: Sage.

Moy, P., Manosevitch, E., Stamm, K., & Dunsmore, K. (2005). Linking dimensions of Internet use and civic engagement. Journalism and Mass Communication Quarterly, 82(3), 571– 586.

Musso, J., Weare, C., & Hale, M. (2000). Designing web technology for local governance reform: Good management or good democracy? Political Communication, 17(1), 1–19.

Norris, P. (2001). Digital Divide? Civic Engagement, Information Poverty and the Internet Worldwide. Cambridge: Cambridge University Press

Norris, P. (2003). Deepening democracy via e-governance. Retrieved February 25, 2008,from:http://ksghome.harvard.edu/_pnorris.shorenstein.ksg/Acrobat/ World%20Public%20Sector %20Report.pdf.

Shah, D. V. (1998). Civic engagement, interpersonal trust, and television use: An individuallevel assessment of social capital. Political Psychology, 19(3), 469–496.

Scott, J. K. (2006). ''E'' the people: Do U.S. municipal government websites support public involvement? Public Administration Review, 66(3), 341–353.

Sociopolitical Aspects of Electronic Voting

Vanja Malidžan
Singidunum University, Belgrade
Full address: I. G. Kovačića 13, Banja Luka, Bosnia and Herzegovina
E-mail: vanja.malidzan@gmail.com

Summary

There are a number of countries throughout the world that implemented some type of electronic voting in the past three decades. In this paper we are examining the experiences from these countries based on the conducted social researches and the political effects that electronic voting brought. From the social point of view it is evident that voters are not so excited to switch from traditional, sociocentric voting to faceless electronic voting. Looking from the political point of view, arguments that another voting channel will increase political participation do not hold so strong against the numbers from elections where electronic voting was one of the options. Nevertheless, with the growing infiltration of information and communication technologies in our society, it is almost certain that all countries, including Bosnia and Herzegovina and countries from the Balkans region, will implement some kind of electronic voting in near future. That's why we are trying to give several directions how to implement electronic voting as painless as possible in this paper.

Key words: electronic voting, elections, remote voting, sociopolitical aspect

Introduction

In the past fifteen years we witnessed exponential growth of internet use. Information and communication technologies (ICT) became an integral part of our lives. Ordinary citizens, companies, non-governmental organizations and governments use internet and computers more and more for their daily operations. We are using internet for e-banking, e-shopping, we can apply for some documents and find endless amount of information there.

Simultaneously with the development of ICT, modern technologies find its role in political life. Political parties and candidates are running their campaigns on internet, doing fundraising, lobbying, collecting signatures for the petitions and recruiting new members. The power of internet to improve communication and access to information suggests its increasingly important relationship with the electoral politics. ICT will likely continue to have a significant impact on the nature of democracy in countries worldwide. One of the development directions is usage of modern technologies in electoral processes.

In last 30 years many countries throughout the world experimented with the different types of electronic voting with more or less success. While electronic voting promises significant improvements compared to traditional voting, there are certain risks associated with it. Most of these risks are related to public confidence in the security of the voting process and social nature of elections.

All risks and benefits should be carefully analyzed prior to any attempt of electronic voting implementation in one jurisdiction. We are going to present some experiences and research results from the countries that tested or fully implemented electronic voting. These results help us derive some recommendations how to painlessly implement electronic voting in countries like Bosnia and Herzegovina which neither tested nor politically discussed electronic voting of any form so far.

In this paper we are focusing on sociopolitical aspects of electronic voting, but there are other aspects that need to be considered as we are going to see later.

Definition of Electronic Voting

From the first use of now widely accepted 'Australian ballot' in mid 19th century, people were thinking how to use machines to improve electoral procedures. America was a leader, introducing mechanical lever machines, punched cards for voting and optical scanners.[1] First use of electronic machines in real elections goes back to 1970s in America. Since then, many electronic voting system prototypes have been deployed all over the world.

Today, when we say electronic voting (e-voting, eVoting), we assume use of some electronic means in some or all voting procedures. We can distinguish two main types of e-voting:

- Direct recording electronic (DRE) voting machines; and
- Internet voting.

DREs are machines (computers) located at the polling stations, where voters can cast their votes using the touchscreen or some type of simplified keyboard. Depending on the legislation, vote counting is carried out almost instantly at the polling station after the polls close, or the electronic ballots are transferred on memory modules to the central location where counting should take place.

Internet voting utilizes the public communication infrastructure for the procedures of voter's authentication and authorization, ballot casting and verification. Internet voting systems can be grouped into three general categories (Oostveen, 2007):

- **Poll site** internet voting refers to the casting of ballots at the public sites where election officials control the voting platform;
- **Kiosk** voting located in the convenient public places like community centers, libraries, post-offices, train stations or schools; and

---

[1] A Brief Illustrated History of Voting, The University Of Iowa Department Of Computer Science, http://www.cs.uiowa.edu/~jones/voting/pictures/

- **Remote** internet voting refers to the casting of ballots from any computer connected to the internet from anywhere in the world.

The author of this paper has strong attitude against any investment in the research and development of DREs in countries without legacy systems of that type. With the exponential growth of internet use, spending money and time on voting machines would be the waste of valuable resources. Online remote voting should be considered an essential convenience in modern society (Madise and Martens, 2006 in Goodman et al., 2010). That's why in this paper we are focusing on the internet voting, namely remote internet voting.

Aspects of Electronic Voting

Voting is the fundamental cornerstone of modern democracy which cannot be conceived without elections. It is the main mechanism to disclose the opinion of a group (of voters) about an issue that is under consideration. Since elections are so delicate and important in modern democratic countries, voting procedures have to be carefully examined in all belonging aspects. Same holds for the electronic voting. There is an ample of aspects that need to be considered:

- technical,
- legal,
- political, and
- social.
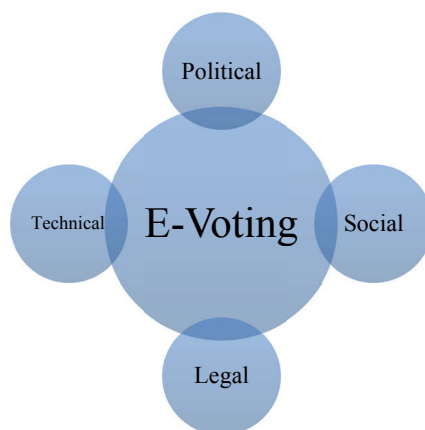
Figure 1. Aspects of Electronic Voting

Very often researches depart from technological characteristic of new technology, neglecting all other aspects. This is true not only for e-voting, but many other implementations of the modern technologies in everyday life.

More recently, it has been recognized that an important characteristic of modern technology is the existence of complex and large technical systems - spatially

extended and functionally integrated socio-technical networks (Mayntz and Hughes, 1988 in Oostveen, 2007) in which research of the 'social' and 'technical' side cannot be separated. As Lamb et al. (2000) point out: "ICTs do not exist in social or technical isolation".

Large technical systems have following properties (Oostveen, 2007):

- They affect many people and institutions;
- They are complex: political, legal, administrative, organizational and technical issues are relevant in the design, development, implementation, maintenance, and use of these systems;
- They are infrastructures, and face difficult issues of standardization;
- They generally embody political ideas and ideologies.

In the rest of the paper we will try to give advantages that e-voting is bringing, but also risks associated with it.

Advantages of Electronic Voting

In the past 25 years organized participation in society gradually has declined in many countries of Europe. (Van Dijk, 2010). Electronic voting is seen as a mean to increase participation, notably political participation. Key argument for the advocates of e-voting is increased convenience brought by the new technology. E-voting has the potential to make the voting process **easier and more accessible** for electors. There is a potential to eliminate long line-ups at the polling stations and better address accessibility issues for people with disabilities, those suffering from illness, those serving in the military or living abroad, those away on personal travel, and other groups of citizens such as single parents who may find it difficult to visit a traditional polling station.

With regard to **young people** aged 18 to 30, internet voting may be the way to engage them in the elections. This group of voters seems to be the hardest to reach by traditional political methods. E-voting may use their natural interest and familiarity with modern technologies.

Internet voting allows **greater secrecy** for electors with disabilities (including visually or hearing impaired). By voting electronically and therefore unassisted, these voters are afforded a greater degree of anonymity when casting a ballot.

By the nature of its work and life style, modern man & woman have less space constraints. Krzywiecki and Kutyłowski (2010) think that the traditional voting attached to the place of living is surpassed model of 19th century. E-voting is offering **voting without geographical and time constraints**.

Another promise of electronic voting is **the universal verifiability** which is impossible with the traditional voting. In traditional voting we have to trust our proxies (poll watchers) which observe voting procedures directly at the polling stations. In e-voting any interested party using internet can check if only eligible voters voted, if each voter casted only one ballot and if all casted ballots are included during tallying.

160

E-voting **does not have a 'point of no return'** after which voter cannot change his/her mind. That point in traditional voting is the moment when voter puts the ballot into the ballot box. Current implementations of e-voting offer a possibility to cast a ballot number of times, and only the last one will be the one that counts.

From the political point of view, e-voting has the advantage of **faster tallying** and **reduced number of invalid ballots**. Reduced time gap between elections and the announcement of final results decreases uncertainty of electorate and thus restore the trust in political system. Same applies for lower number of invalid ballots.[2]

Over the long term all types of internet voting have the potential to be **less expensive** to operate and execute than traditional paper ballots which require setting up and staffing polls. However, the start-up costs can be very high.

Electronic Voting Risks

Critics of internet voting express concern about security and about the lack of equal access to the internet for all citizens.

If we concentrate on social perception of electronic voting, one thing that needs to be considered is **technophobia** - fear or dislike of advanced technology or complex devices, especially computers. Although this fear could look like a little bit irrational, it is evident and has to be considered.

Another reason against e-voting is the **importance of elections for local community**. For the large number of citizens a moment when they cast their ballot is the symbol of their political activity. Having in mind the fast modern life style, traditional voting is one of the rare moments when people can meet and talk with their neighbors.

The **loss of the civic ritual** is also mentioned in many articles about internet voting. E-voting would make elections less of a community event, which might create a greater gap between citizens and the government, thereby decreasing political participation. Some believe that voting is more than the simple act of disclosing one's political preference; it is a vital public ritual that increases the social cohesion and unites citizens together.

Important risk is a potential for '**digital divide**'. This divide can occur in several different ways. There is a digital divide between those with and without computers, with and without (fast) internet access. Technical opportunities are not the same for wealthy and poor people, those living in urban and rural areas. Basically, internet voting has the potential to create divides with respect to many socio-economic variables: income, education, gender, geography, race and ethnicity.

---

[2] But e-voting system **must have** an option for voters to cast invalid or blank ballot, since it is a common way for voters to protest against offered choices.

Internet voting presents greater opportunity for **frauds, bribery and coercion**. Voting from home enables family voting. Also, it is possible that coercer will vote on someone's behalf, since voting is happening at the remote private place. That's one of the greatest risks of remote voting compared to the traditional voting in controlled environment of the polling station.

There is an additional opportunity for fraud or accidental failure in e-voting systems if voter's computer is attacked by viruses. A potential weakness of internet voting is its vulnerability to a variety of **hacker created problems** (web site spoofing, denial of service attacks, etc.). Individual hackers, criminals, and foreign intelligence services are among those who might try to manipulate the vote or destroy the technology used to run the election. (Rogerson, 2003).

Last but not the least, traditional voting procedures are rather simple and effective and many will argue that there is **no need for electronic voting** at all.

Implementation of Electronic Voting in European Countries

Some countries employ electronic voting for years and other countries begun with the testing and pilot projects. Yet in some cases electronic voting has been abandoned even before it was tested, no matter how much money or time was spent. Reasons are varying but in general it's because public concerns were raised about security, reliability and privacy.

Netherlands, one of the European countries with the longest history of electronic voting, said no to e-voting in 2008. The Ministry of the Interior decided that way after reviewing extensive research which indicated that none of the available machines offered adequate privacy and security safeguards.

In Germany, Federal Constitutional Court decided in 2009 that electronic voting used for the last 10 years, was unconstitutional and therefore not to be used for the upcoming elections. The court ruled that the use of the electronic machines contradicts the public nature of elections.

Irish Minister for the Environment announced in 2009 that the electronic voting system was to be suspended, due to the cost and the public's dissatisfaction with the current system after spending more than €54 million on the machines.[3]

However, there are successful stories with the electronic voting. Estonia is the only country which offers its voters possibility to vote by internet in absolutely binding national elections. Estonia has been using internet voting since 2005 in five election cycles so far.

In Norway it has been announced that electronic voting over the internet will be tried out in certain areas for local elections in September 2011, with the ultimate

---

[3] Irishtimes.com, http://www.irishtimes.com/newspaper/breaking/2010/1006/breaking26.html

goal of implementing full general availability for internet voting for the 2017 parliamentary elections.[4]

Since 1998, the Swiss government has actively pursued the implementation of electronic voting in its elections. Successful systems have worked in the test cantons of Geneva and Zurich (Gerlach and Gasser, 2009).

Existing Researches

Across the world a number of interesting sociopolitical researches have been carried out on the topic of electronic voting. Most of them followed binding elections or test projects that used e-voting. We are going to present selected findings from these researches.

Reniu (2009) carried out different surveys in Spain, Mexico and Argentina, trying to outline which are the main perceptions of citizens when using different e-voting solutions. This research covered public and private e-voting events being these events binding or not, using only e-vote or together with the traditional vote, and with remote voting or DRE.

Surveyed people felt **quite satisfied** with the use of e-voting, rating their satisfaction with a median value of 4,2 out of 5. Another important finding is the **lack of trust in e-voting**, especially binding remote internet voting. Some of the examined projects had an option for the voters to choose between e-voting and traditional voting. When asked why you chose to vote traditionally, common answer was: "We always cast our votes using paper ballots and transparent urns and, more important, we meet each other at the polling station, sharing the democratic liturgy."

Chart 1. Reasons for opting traditional voting (Reniu, 2009)

---
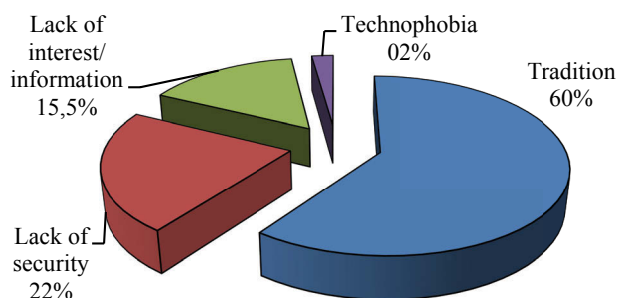
[4] Prosjektdirektiv for e-valg 2011, Kommunal- og Regionaldepartementet. February, 11 2009. http://www.regjeringen.no/upload/KRD/Vedlegg/KOMM/Evalg/Prosjektdirektiv_evalg2011_v10 1.pdf

Same research confirms general support for the remote voting (almost 80% of respondents) but with the condition to have an option to vote traditionally or not.

Oostveen (2007) conducted researches in France, England, Italy and Finland. E-voting systems that were analyzed included a variety of e-voting technologies: a smart-card internet voting system which was used at home, work, and in school; a web-based voting system; and voting computers at the polling booth.

Among other findings, one that is attention-grabbing is **lack of trust in privacy** of remote internet voting system (TrueVote[®]).

Table 1. Opinions about usability of the remote voting system (Oostveen, 2007)

	Yes	Neutral	No
Trust in security (against fraud and hackers)	60%	17%	23%
Trust in secrecy (privacy)	5%	11%	84%
Trust in accountability (verify the vote)	62%	16%	22%
TrueVote is easy to use	92%	4%	4%
TrueVote is fast	77%	13%	10%
TrueVote is easy to install	65%	15%	20%
TrueVote is robust (not vulnerable for pincode/pincard loss)	52%	21%	27%

While voters believe in the security of e-voting system, 'Big Brother' effect is over-emphasized and almost no voters believe that their ballot casted over the internet will remain secret. There is a fear that the vote is not anonymous since the operations would be registered on the computer. Many of the participants believe that governments have a lot of information about citizens and are afraid that e-voting will only add to that information.

Elections Canada survey data (Goodman et al., 2010) offers insights regarding the probability that people will vote by internet. In 2000, for example, 47% of the respondents report being likely to vote online, in 2008, **interest rose** to 54% of the electorate. Aside from those voters over the age of 54, a majority of respondents in all age groups indicate that they would be **likely to make use of the online voting** if the service was available.

Based on the research results from three Canadian pilot projects with the internet voting for local elections, we can conclude that implementation of internet voting as an additional option **does not automatically increase the turnout**. Interesting data comes from Canadian city of Peterborough - 70% of online voters were **45 and older**, and the highest rate of use was among electors aged 55 to 64. Only 14% of those aged 18 to 34 voted online (Goodman et al., 2010). The higher usage rate among 'baby boomers' is interesting because most survey data indicates that young people are more inclined to report using, or saying they would make use of, internet voting than other cohorts of electors.

Experiments with internet voting in Switzerland can be considered successful. Switzerland is one of the most developed countries with high rates of internet penetration. Also, Switzerland has a long tradition of postal voting, and remote

internet voting is seen as an extension to this popular method. Data discloses that there is no digital divide with the respect to education or gender, but one is visible in terms of **age and internet competence**.

One of the key reasons for success of internet voting in Switzerland is trust in state's voting system, correct outcome of elections and **the trust in country's political system** in general (Gerlach and Gasser, 2009). Although e-voting system in Switzerland is technically far from being 'state of the art', nobody even raised a question about secrecy nor tried to fool the system.

Estonia is the only country in the world to have legislated internet access as a social right (Trechsel, 2007 in Goodman et al., 2010) and is one of the most electronically enabled countries in Europe. These elements, and the fact that internet voting in Estonia can be considered an electoral success, made it an important case for the research.

This year the turnout in Estonia further increased in comparison to the years before and especially to the record turnout in 2009. In total 140.846 Estonians casted their vote over the internet.

Chart 2. Turnout and number of internet voters in Estonia[5]
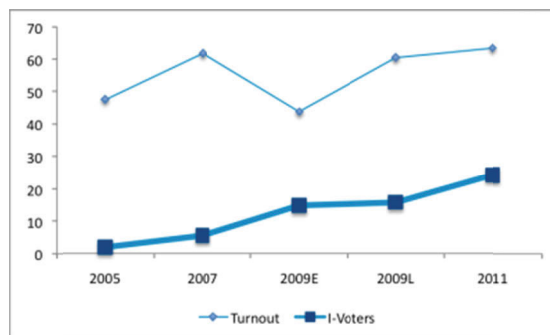
The number of young voters is constantly at 10% of the internet voters. In contrast, voters above 55 account for 18% of the internet votes on average supporting underlining the trend of the '**silver surfer community**' (similar to Canadian case mentioned earlier). This may lead to the conclusion that the hypothesis about age usage may be doubtful.

Conclusion

Lessons we can learn from countries that use e-voting helps countries like Bosnia and Herzegovina to draw its own road map for the implementation of e-voting systems. The blank paper we have in front of us may be our advantage in sketching trouble-free path to the e-voting. We can avoid repeating failures that

---

[5] Source: e-voting.cc

165

others had. Here are some best practices and recommendations for the implementation of e-voting systems:

- Politics heavily rely on interpersonal relations and cannot be entirely transferred to cyber world, since people still want some community gatherings;
- Citizens will have trust in e-voting system only if they have trust in the political system. Both needs to be built simultaneously;
- E-voting has to be complementary tool, not sole way for voting;
- If properly implemented, e-voting system can attract all age groups of voters;
- Start with small-scale tests and gradually increase the number of included voters. Avoid too many non-binding test projects;
- All involved actors (voters, political parties and candidates, media, NGOs etc.) need to be involved in the development process;
- Education of voters is crucial. Do not assume that voters will know how to use electronic systems.

There is almost no doubt that internet voting will progressively be implemented in most democratic countries. And, for the countries like Bosnia where no notion about e-voting exists, it is better to skip the step with electronic voting machines and go directly to the internet voting. Development of such systems can be hard and expensive, so buying some proven system sounds like a reasonable choice, with the necessary adjustments to specific legislation.

References

Gerlach, J; Gasser, U. Three Case Studies from Switzerland: E-Voting. *Berkman Center for Internet & Society at Harvard University*, publication No. 2009-03.1

Goodman, N; Pammett, J.H; DeBardeleben, J. A Comparative Assessment of Electronic Voting. *Strategic Knowledge Cluster Canada-Europe Transatlantic Dialogue for Elections Canada*, 2010

Krzywiecki, L., Kutylowski, M. Lagrangian E-Voting: Strong Privacy and Verifiability on Demand. *TRUST 2010*, LNCS 6101, Springer Verlag 2010, pp. 109-123.

Lamb, R., Sawyer, S. and Kling, R. A Social Informatics Perspective on Socio-Technical Networks. In Chung, M.H., ed. *Proceedings of Americas Conference on Information Systems* Long Beach, CA, 2000, pp. 1613-1617.

Oostveen, A. M. Context Matters: A Social Informatics Perspective on the Design and Implications of Large-Scale e-Government Systems. PhD Thesis, *University of Amsterdam*, 2007

Reniu, J. M. E-voting: An Analysis of Sociopolitical Acceptance. 21st IPSA World Congress, Chile, 2009

Rogerson, S. Cyber terrorism and the threat to democracy. originally published as ETHIcol in the *IMIS Journal* Vol. 13 (August 2003), No. 4

Van Dijk, J. Study on the Social Impact of ICT - Topic Report 3 (D7.2), *Universität Siegen, Fachbereich Wirtschaftsinformatik und Neue Medien, Germany*, April 30, 2010

Mobile Parking Payment Service

Aleksandar Skendžić
University of Applied Science
Bana Ivana Karlovića 16, Gospić, Croatia
askendzic@velegs-nikolatesla.hr

Božidar Kovačić
Department of Informatics, University of Rijeka
Omladinska 14, Rijeka, Croatia
bkovacic@inf.uniri.hr

Summary

Electronic business (e-*business), as a phenomenon popularised by the Internet and the development of networking technologies, is entering a period of rapid expansion in the field of portable communications devices and their respective applications.*
In the present day there are numerous services available on the Internet via wireless networking technologies. Mobile phones have proved themselves to be an ideal platform for conducting e-business because of their practicality, availability and simplicity of usage. One of the examples of practical efficiency of a mobile phone is the option of paying for services via short textual messages (SMS) through Wireless Application Protocol (WAP). The SMS parking service is a relatively fresh service on the market (not the only one of the sort). It is used primarily in the urban, city areas world wide and it represents a significant contribution to e-business. Simplicity, compatibility, availability and the option of paying for a service at any time from any place makes this way of conducting payment very acceptable to the end users. Besides its practicality and simplicity mobile payment is widely available. There is a report available at any point on the current state of the charge as well as on the expiration time of the service. The WAP protocol and the SMS service represent the foundation of concepts on which the electronic charge of services through portable communications devices is based. Using alternative methods of the mobile payment service like integrated applications on the portable communications device, using Near-Field Communications (NFC) technology or the Internet requires more time and larger amounts of transferred data which finally means high price of the service.

Key words: e-business, GSM, World Wide Web, WAP, SMS, Voice, HTML, mobile payment, mobile business.

Introduction

Intensive expansion of the Internet together with the networking of business companies lays a good foundation for further development of e-business. E-business is developing intensively therefore in certain areas of business conduction it plays a very significant role. Development of e-business brought in major changes in ways that profitable and non-profitable organizations conduct business. Some of the relevant characteristics which contributed to the importance of e-business are: ability to transfer large amounts of data across great distances, simple updates, availability, direct payment via the Internet, usage of advanced network infrastructures for specific services like the mobile phone payment of the parking service, Online banking, buying tickets for public transportation using a mobile phone and many other advanced services. According to ESPRIT-[1] (a project of the European Union) the following definition for e-business is used: "*Electronic business is a general concept which represents all forms of business transactions or exchange of data which is performed with the help of Information and communications technology (ICT). It refers to the business transactions and data exchange among companies, between companies and their customers or between companies and public administration. Electronic business also includes the electronic market of goods and services.*" One of the new forms of conducting business is the mobile business which primarily relies on the Global System for Mobile Communications (GSM). Mobile business represents one of the newer business models. Figure I. describes mobile business. Figure II. describes mobile payment through today's popular technologies such as WAP, SMS, Internet and integrated mobile phone applications. Figure III. shows SMS parking service conceptual scheme in theory and figure IV. gives an example of "m-parking" service in the city of Zagreb.

Mobile business

Mobile business is a part of electronic business based on the application of specific mobile technologies which give companies a market advantage in the mobility of workers and customers. This means that the workers are neither tied down to their offices nor to their houses. It also means that the consumers can order and pay for a product or a service from any place nation or world-wide with a single restriction - the signal coverage of their mobile operator. In the mobile business domain the following areas can be differentiated:

- mobile business communication;
- mobile marketing;
- mobile payment.

---

[1] European Strategic Program on Research in Information Technology: http://www2.cordis.lu/esprit/src/ecomproj.htm

Mobile technologies used in business purposes can be found mostly in the area of business communications which imply the exchange of SMS messages between the providers of the service and the consumers. Mobile communication among the employees is established via Intranet. According to a research conducted by the company Forrester Research[2] the most significant obstacles on the way towards development of mobile business are:

- concerns about the safety of transactions;
- possible previous bad experiences;
- ignorance about proper usage of technology.

GSM is the most widespread standard for mobile telephony used in mobile business. Mobile business represents the application of mobile technologies in business purposes, providing services, marketing and payment in order to improve business efficiency. Efficiency refers to lower expenses and to competitive advantage on the market. With the new, third generation of mobile phones (3G) which support better bandwidth a new palette of services became possible. The services can be classified into three categories:

- network services, which represent telephony through network administration;
- additional services related to the network like SMS, voice mail etc.;
- services related to a third party like mobile payment involving companies and banks.

Mobile payment

Mobile payments include all payments conducted with the help of a mobile phone, laptop or any other portable electronic device through network technologies and the Internet. Mobile payment can be conducted through the following technologies:

A) WAP protocol for data exchange in mobile networks

Wireless Application Protocol (WAP) enables access to internet services and content exclusively adapted for display on mobile phones. WAP has the purpose of transferring information so that it is adjusted to mobile users taking into concern the limited capabilities of data transfer through mobile telephony radio channels. It uses different forms of data compression in order to minimize the number of bits transferred through the wireless media. WAP applications include email, electronic marketing, online banking, transactions on the electronic market, information about phone calls, services, unified messaging, weather and traffic reports, news, information about sports, online file storage etc. WAP relies on Wireless Markup Language (WML). WML is a programming language optimised for the wireless transfer of data. The transfer

---

[2] Forrester Research, url: http://www.forrester.com/rb/research

of data is based on the Wireless Transport Protocol (WTP) standard. WML shows the classical internet pages in a format which a mobile telephone can recognise and display on its screen. WAP as a communication language which enables mobile phones access to Internet services uses advanced wireless technologies of data transfer. Taking into account the speed of data transfer, mobile payment of services is preformed through the following data transfer technologies: GPRS[3], EDGE[4], UMTS[5] and HDSPA/ HSUPA [6].

B) Short message service
Mobile payment for services is also possible through the SMS service which is primarily used for textual communication on mobile phones. The SMS service means sending short textual messages through GSM standard for mobile telephony. The incoming and outgoing SMS message can contain maximally 160 alphanumeric characters. On its way from a sender to a receiver, the SMS message goes through the so called SMS centre, which is in reality a computer functioning as a server. The server has the function of distributing SMS messages to the end users and also ensures confirmation messages containing data about successful deliveries or delivery failures to senders.

C) Integrated applications for mobile payment adapted for usage on mobile phones
Mobile payment for services is also possible through internet applications and networks that are using Hyper Text Transfer Protocol (HTTP) and World Wide Web (WWW) service. A good example is the new version of a mobile operating system Android[7] (current version *2.3*[8]) which among its other features has integrated backup for mobile payment through Near-Field Communications (NFC). NFC is a chip that enables communication between devices within small distances. It makes paying bills possible using just a mobile phone and the appropriate terminal.

---

[3] **GPRS** (General Packet Radio Service) enables data transfer and mobile access to the Internet, bandwidth up to 80 kbps.

[4] **EDGE** (Enhanced Data rates for Global Evolution) is a third generation (3G) mobile networking technology, enables data transfer and mobile access to the Internet, bandwidth up to 220 kbps.

[5] **UMTS** (Universal Mobile Telecommunications System) is a third generation mobile networking technology, enables data transfer and mobile access to the Internet in Croatia, bandwidth up to 384 kbps.

[6] **HSDPA** (High Speed Downlink Packet Access) and **HSUPA** (High Speed Uplink Packet Access) are upgrade to the UMTS network, enables data transfer and mobile access to the Internet in Croatia, bandwidth up to 7,2 Mbps.

[7] **Android** is an operation system specially adapted for use on mobile phones and other mobile devices like tablet computers.

[8] Current date: July 6, 2011.

D) Using laptops with mobile access to the internet network
Paying for services by using the Internet, the HTTP protocol and the WWW service is based on using Internet applications whose sole purpose is paying for services, for example PayPal. Taking into consideration the popularisation of the so called "smartphone" devices, which have reached a population of 600 million users world-wide with the growth tendency towards 1 billion by the year 2015, this way of paying for services, might compete and finally bring in new standards to the mobile payment service. In March 2011 Berg Insight reported that the global shipments of smartphone devices increased 74% in the period of 2009 – 2010.

Paying for parking using a mobile phone

In the present day there are many solutions in use for payment using a mobile phone. One of the main differences between them is the value added carrier which enables the payment service and also determines the way in which the service can be used. The carrier of the payment service by using a mobile phone can be:

- mobile operator (Vipnet, T-mobile, Tele2);
- the parking organiser (concessionaire) with his own solutions;
- banks which conduct business with the citizens (current accounts, credit cards);
- other companies specialised in paying with mobile phones.

Paying for a parking service by mobile phone can be made possible through:
- using a mobile phone – WAP;
- using a mobile phone – VOICE;
- using a mobile phone – SMS.

A) Advantages of paying for parking by mobile phone for users:
- Using the service without special need for check-ins or registrations and simpler and faster paying for parking;
- "Remote paying" – the user doesn't have to be in the vicinity of the vehicle in order to pay for parking or to extend his parking time. The user can pay for parking for any vehicle at any location (where the service has been introduced);
- "Reminder" – reminder about the expiration of paid parking time comes in 5-10 minutes before the expiration of the paid parking time;
- Availability of service - the service is available 24/7 which enables the user to pay for parking in advance.

B) *Advantages for the concessionaires:*
- Additional way of paying for parking – introducing this option of payment will make many drivers pay for parking more frequently and on time;
- Payment on time will decrease the number of parking tickets and by that the expenses of the concessionaires will drop;
- Decrease of operative expenses of the concessionaires.
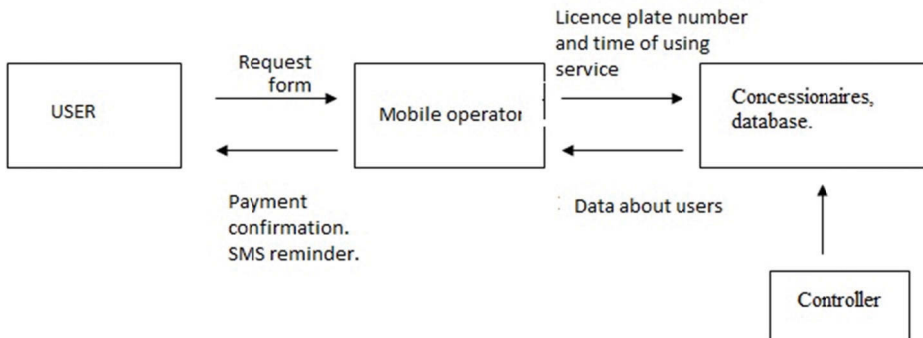
Figure 1 - A Functional schema of an SMS Parking service

Figure 1 shows a functional schema of the SMS parking service. That means that user is sending request from a mobile phone using SMS service. SMS includes the exact form: car registration number like GS891BH (uppercase together connected). SMS is sending to a mobile operator number which forwards information to the database of the mobile parking concessionaire. I the same time, concessionaire server adds car registration number into the database together with other information such as start and end time using parking service. In the central concessionaire server, data is processed and put into separate databases for each parking operator. The data is then sent to the parking operators' servers so they can deal with any possible complaints from drivers. Also, on the basis of data obtained from the central concessionaire server, the parking operator is able to issue an invoice for all payments made through the mobile parking payment service. Table 1 gives technologies comparison for mobile payment.

Table 1. Comparison of technologies for mobile payment

Payment service	Advantage(+)	Disadvantage (-)
Short Message Service (SMS)	Compatibility of prices	Limited to 160 characters. The price of sending SMS message.
Integrated applications for mobile payment	Simplicity of usage, optimization for mobile phones	Compatibility with mobile operating systems, availability, price for using data transfer
WAP – pages adapted to the WAP protocol	Access speed, integration into the mobile phone menu, optimization for mobile phones	Limited display on the screen, price for using data transfer
Internet payment services (PayPal, e-gold, AlertPay...)	Total control of the user account (transaction times, services used...)	Demands an open bank account with deposited funding, obligatory registration, adapted for laptops and 'smartphone' devices

Charging for parking with the"m-parking" service

Most cities in the world have problems with the growing need for parking places. Segments of streets, which are generally public property, are used for parking and make living in the cities bearable. The development of the wireless communication systems technology and its application in everyday life enabled us to use mobile communications to pay for parking. Croatia started the charging parking by mobile communications service called "m-parking" in its capital Zagreb. The pilot project was arranged with only one mobile operator. Afterwards The Croatian Parking Association (CPA) developed a unified, unique system of payment and control of payment in all major towns in Croatia which had "m-parking", connecting the existing mobile operators into a system. Rapid development and application of mobile telephony created prerequisites for the application of modern technologies into the system of payment and control of parking. The penetration of mobile telephony in the year 2008 was approximately 75%[9].

Figure 2 shows a conceptual schema of the "m-parking" system – paying for parking by a mobile phone. Key steps in the process of charging for parking in "m-parking" are the following: processing requests, control over the parking payments and transfer of data between the concessionaires. The whole system is based on data exchange between mobile operators' servers, central server of the CPA and the servers of the concessionaires.

---

[9] Source: http://www.t.ht.hr/press/novosti_disp.asp?id=1372 (accessed 14.4.2011.)
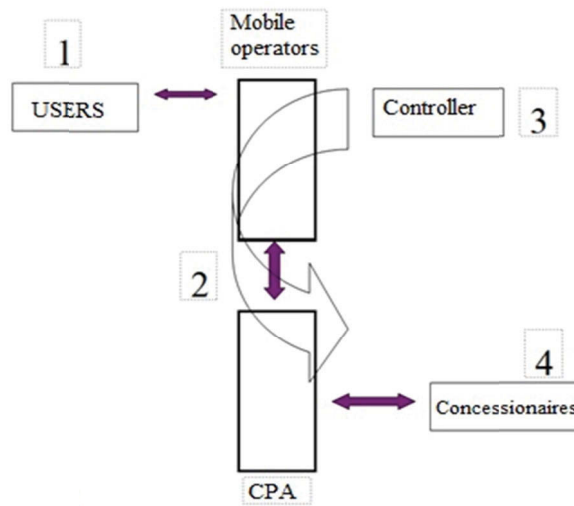
Figure 2. Schema of a conceptual solution to the m-parking service[10]

1-2 Requests processing
The central server of the CPA is connected to servers of the mobile operators. Servers of mobile operators forward SMS messages (requests) to the CPA server, which according to Regulations about payment and control of parking for individual towns sends confirmations (prices for parking along with the expiration time of the paid parking services).

3 *Parking payment control*
The central server of the CPA creates databases on a server assigned to provide data about the completed m-parking payments so that the controllers on the field can perform quarries about statuses of individual vehicles. A group of controllers for every concessionaire has access only to specific data which refer to payment for parking by mobile phones in their respective town.

4 *Data transfer: CPA – concessionaires*
In the central server of the CPA data is processed and included into separate databases for each individual concessionaire – member of the CPA. Data is then sent to concessionaires' servers so they can deal with the immediate complaints of the drivers. Likewise, based on data received from the central CPA server, the concessionaire is able to deliver receipts for all payments executed through the m-parking service.

---

[10] Source: http://infoartgroup.com/wp-content/uploads/mParking-presentation.pdf

Conclusion

Popularisation of "smartphone" mobile devices opens doors to new possibilities in the mobile business development. The work on matching new technologies in practical use is intensive. The figures - 600 million "smartphone" users with the growth tendency towards 1 billion by the year 2015 speak for itself. The SMS payment service is introduced de facto as a standard in cases like paying for services like parking. One of the reasons lies in the simplicity of usage and lower expenses, but also in insuring compatibility with the mobile devices. "M-parking" service has lots of advantages for consumers but also have two main disadvantages. First, the price of sending an SMS message that has been determined in accordance with valid price lists of mobile operators (received messages are free of charge). Second, one SMS message only covers one parking hour. Using alternative methods of mobile payment service like integrated applications on mobile devices, NFC microchips, the Internet requires more time and transfer of larger amounts of data which finally means higher price of the service.

References

Berg Insight http://www.berginsight.com/ShowReport.aspx?m_m=3&id=114 [10.03.2011]

Čerić,V., Varga,M., Birolla, H.: Poslovno računarstvo. Znak, Zagreb, 1998. str. 264-304.

Croatian Parking Association (CPA): http://www.rao.hr [10.03.2011]

ESPRIT: http://cordis.europa.eu/esprit/home.html [08.03.2011]

Forrester Research: http://www.forrester.com/rb [10.03.2011]

Hrvatski telekom: http://www.t.ht.hr/press/novosti_disp.asp?id=1372 [10.03.2011]

Infoart grupa: http://infoartgroup.com/msolutions/mparking   [12.03.2011]

International Telecomunication Union, http://www.itu.int/newsroom/press_releases/2010/06.html [10.03.2011]

International Telecomunication Union, http://www.itu.inte [10.03.2011]

Spremić, M.: Računalna mobilna prodaja // Informatorova birotika i informatika, br. 4799, ožujak, 2000. str. 7-12.

Šehanović, J., Hutinski, Ž., Žugaj, M..: Informatika za ekonomiste. Sveučilište u Rijeci, Pula, 2002.

Spremić, M.: Elektroničko poslovanje i stvaranje nove vrijednosti // Suvremeno poduzetništvo, br. 6/2003., str. 10.

Vipnet: http://www.vipnet.hr/poslovni-korisnici/mobilni-internet/ostale-podatkovne-usluge/o tehnologijama-i-pokrivenosti [10.03.2011]

Evolution of Collaboration by Using Microsoft Office Applications

Ivica Buzov
Administrative High-School
Varšavska 17, Zagreb, Croatia
ibuzov@gmail.com

Summary

Collaboration, especially the team work one, and implementation of projects are common in business environment nowadays the same way are the collaboration possibilities of mostly used office applications.

The author of this article analyzes possibilities of collaboration by using Microsoft Office's business applications: Word, Excel, PowerPoint, and Access. The versions of Microsoft Office 2003, Microsoft Office 2007, Microsoft Office 2010 with Web Apps are included in this article. We can monitor the options mostly from offline collaboration by using Microsoft Office 2003, through the combination of online and offline collaboration, by using Office version 2007 to mostly online collaboration, by using Office version 2010 and especially by using Web Apps and Microsoft Office365 at Cloud.

From each version of Microsoft Office to another one, it is possible to notice the evolution up to online synchronous collaboration options. Some of these features are: comments, tracking changes, document workspace, comparing and combining documents, publishing and sending documents, linking to other sources, collecting data through e-mail etc.

Moreover, we can notice collaboration as being asynchronous or synchronous one, or either client-based (desktop) or server-based (the Internet), extended by using other Microsoft business technologies like SharePoint sites and Groove workspaces with additional communication and collaboration tools.

Furthermore, what is very important in collaboration activities is the security of documents (files) and information, distribution channels and permission levels. This area is also covered, but only in basics.

Key words: evolution, collaboration, Microsoft, Office, 2003, 2007, 2010, Office 365, cloud, Word, Excel, Access, PowerPoint, Groove, SharePoint

Microsoft Office
Microsoft Office Versions, Editions, and Applications
Microsoft Office is widely spread office applications suite in our time. The Microsoft Office system versions are Microsoft Office 2003, 2007, and 2010.[1] There are different editions of Office suites in same version and they include some of the following basic business applications: Word, Excel. Access, PowerPoint, Outlook, or some following additional applications: Publisher, Project, Visio etc. No single edition includes all of these programs. Some programs are available separately.

The Microsoft Office 2010 system includes edition with Office Web Apps on the Microsoft cloud.[2]

New Microsoft Office Applications
So, while Microsoft Office 2003 introduces new applications: InfoPath 2003 and OneNote 2003, Microsoft Office 2007 introduces server-side tool Groove 2007, replaces FrontPage 2003 with SharePoint Designer 2007, and Microsoft Office 2010 replaces Groove 2007 with SharePoint Workspace and introduces Office Web Apps.

Asynchronous Offline Collaboration
In contemporary business environment teams usually work on documents. In team work there are usually needs to preview your colleagues' or employees' work, to make or suggest some changes, to control level of allowed changes with permissions and passwords, and to combine or compare different document versions.

Our focus will be mostly put on common Office applications like Word, Excel, PowerPoint, and Access. All these applications can directly send documents to

---

[1] More information about Microsoft Office versions, editions, and including applications can be found at:
Behera, Suresh. Office Profesional Editions VS All Other Office 2003 Editions @ VSTO. 18.04.2006. http://weblogs.asp.net/sbehera/archive/2006/04/18/443219.aspx (12.06.2011.); Thurrott, Paul. Microsoft Office 2003 Editions Compared. 06.10.2010. http://www.winsupersite.com/article/product-review/microsoft-office-2003-editions-compared (12.06.2011.); Cox, Joyce; Preppernau, Joan. Microsoft Office PowerPoint 2007 Step by Step. Redmond: Microsoft Press, Online Training Solutions, 2007, p. xxviii-xxix; Microsoft Office suites. 2011. http://office.microsoft.com/en-us/suites/ (20.06.2011.); Compare suites available through Volume Licensing. 2011. http://office.microsoft.com/en-us/buy/office-2010-volume-licensing-suites-comparison-FX101825637.aspx (20.06.2011.); Which suite is right for you? 2011. http://office.microsoft.com/en-us/buy/office-2010-which-suite-is-right-for-you-FX101825640.aspx (20.06.2011.).

[2] Bott, Ed; Siechert, Carl. Microsoft Office 2010 Inside Out. Redmond: Microsoft Press, 2011, p. 19-20. Cox, Joyce; Lambert, Joan; Frye, Curtis. Microsoft Office Professional 2010. Redmond: Microsoft Press, 2011, p. xv.

other collaborators. Microsoft Office 2007 introduces, besides the feature of sending, the feature of publishing documents.[3]

Comments
All Microsoft Office versions have the option to insert comments in documents. Comments are usually used to ask the author, to give some suggestions or explain text or for formatting. Comments can be added in Word, Excel and PowerPoint, and not in Access

In the Word application, comments are placed in balloons in marking area and marked by name of its author/reviewer. Reviewing Pane is used to see comments and to edit or delete it.[4]

In the Excel application, comments are added to cell, and after you add a comment, comment flag appears in the cell, and a comment appears next to the cell. The comment can be shown all the time or hidden and appears when mouse pointer is moved onto the cell that contains comment. It's possible to select every cell containing comment by using 'Go To Special' dialog box.[5]

In the PowerPoint application, comments can be added to slide or slide object. If the slide object such as title or graphic is selected before adding comment, the comment icon appears in the upper-right corner of the object. Otherwise, the comment icon appears in the upper-left corner of the slide. The comment is hidden being clicked away from a comment box. However, there is a small icon with author's initials and number left. These comments can be all hidden, edited, or deleted.[6]

---

[3] There are some differences between sent options at Microsoft Office 2003 applications: Access 2003 has features to be sent directly to the mail recipient and to mail the recipient as an attachment, PowerPoint 2003 has additional features to be sent to the mail recipient for review, to Exchange folder, to online meeting participant, to recipient using Internet fax service, Excel 2003 has additional feature to be sent to routing recipient, while Word has additional feature to be sent to recipient using fax-modem.

Microsoft Office 2007 introduces options to send and to publish. There are two options to send a copy of document to other people using e-mail message to send as an attachment or using Internet fax service to fax the document. There are options to publish document different for Office applications: Excel service (Excel 2007 only), package for CD and publish slides creating a Slide Library of slides (PowerPoint 2007 only), blog as blog post (Word 2007 only), document management server (Word 2007, Excel 2007, and PowerPoint 2007), and create document workspace for Word 2007, Excel 2007, and PowerPoint 2007 files. More: Share and publish your files in the 2007 Office release. 2011. http://office.microsoft.com/en-us/word-help/share-and-publish-your-files-in-the-2007-office-release-HA010214069.aspx?CTT=1 (22.06.2011.)

[4] Cox, Joyce; Preppernau, Joan. Microsoft Office Word 2007 Korak po korak.. Zagreb: Algoritam, 2008, p. 286-289.

[5] Frye, D. Curtis. Microsoft Excel 2007 Step by Step. Redmond: Microsoft Press, 2007, p. 349-351.

[6] Cox, Joyce; Preppernau, Joan. Microsoft Office 2007 Step by Step. Redmond: Microsoft Press, Online Training Solutions, 2007, 176-179.

Comments are useful as notes accompanying documents as suggestions, messages or warnings, but they are not applicable for simultaneous dialogue with others.

Track Changes

Whenever number of coworkers collaborate to produce or edit a document, it is very useful to turn on change tracking, which highlights every change made to the document, but in different color assigned to the user who has made changes. Every user has differently colored change. After the document is finished, changes can be accepted or rejected by using Previewing Pane. It is possible to either preview the changes made by one person only, more users or by the time the changes are made.[7]

There are options to hide all revisions of document, or to show only revisions as deleted text, inserted text, formatted text, or comments. It is possible is to show or hide revisions of one or selected reviewers. Moreover, it is possible to show original version with revisions or without them.[8] Change tracking are used by Word or Excel.

Track changes can be used when multiple authors work on the same document like publishers, lawyers, or businessmen, and it is used during the work on the document. When more then two persons use this option, many different versions and combinations of made changes could be made. Sometimes it is hard to follow which is the most recent and accurate version.

Comparing, Merging, and Combining Documents

In the process of working in a team, or with client, where people involved make certain changes, various versions of the same document may appear. Finally, these documents have either to compare, accept and reject changes, or merge/combine these document versions.

Word 2003 has option to compare and merge document, while Word 2007 has options to compare or to combine documents. Reviewing pane is used in both options, with additional possibility to show both the original document and revised document.[9]

PowerPoint 2003 has option to compare and merge presentations, but this presentation has to be sent as an attachment for a review. For comparing these

---

[7] Frye, D. Curtis. Microsoft Excel 2007 Step by Step. Redmond: Microsoft Press, 2007, p. 351-354.

[8] Cox, Joyce; Preppernau, Joan. Microsoft Office Word 2007 Korak po korak.. Zagreb: Algoritam, 2008, p. 281-285

[9] Cox, Joyce; Preppernau, Joan. Microsoft Office Word 2007 Korak po korak.. Zagreb: Algoritam, 2008, p. 290-292

revisions of presentation, revisions pane is used where there is an option to show selected reviewer's changes.

Comparing, merging, and combining documents are features usually used after finishing work of few persons on same document, but making different versions of this document. It also can be used after receiving different versions of same document. Sometimes it is hard to follow which is the most recent and accurate version.

Asynchronous Network or Online Collaboration

Microsoft Office supports asynchronous online collaboration where documents are usually stored on server, network storage or shared folders and disks and by using shared document workspace, including other server-based business applications and services or products as SharePoint sites or Groove workspaces. All recent Windows versions have option to create shared folder and disks or drives. There can be put files for collaboration with other people, restricted access by using permission level, but only in cases when client PC is turned on and is without document version control.[10]

Shared Workspace

Shared document workspace can be found in all Office versions, but also in some Office applications like Word, Excel, PowerPoint, and Visio.

In Office 2003 it's necessary to assign URL to shared document workspace which contains the copy of shared document. There is an option to invite other people to work on it, to assign tasks, or to link to other resources.[11]

---

[10] Shared disks and shared folders can be used by Windows Explorer or administrative tool Computer Management. Permissions for sharing allows to add or to remove group or user and to set allow/deny i used File Sharing wizard with options to restrict permissions and users to shared folder from permission level Reader to Contributor or Co-owner. Preppernau, Joan; Cox, Joyce. Windows Vista Step by Step. Redmond: Microsoft Press, Online Training Solutions, 2007, p. 250-253.

In order to share files, it is necessary to turn on network discovery and file and printer sharing (they are turned on by default in Windows 7) so computer is visible in the Network group window on a Windows 7 computer. or in the Network folder window on a Windows XP or Windows Vista computer. There are options to share folder with specified user account or groups of users, and to set permission level from default level Read to Read/Write level. On Windows 7 there are also public folders: Public Documents, Public Downloads, Public Music, Public Pictures, and Public Videos. These public folders are turned on by default for Home Network, Work Network, and Public Network connections and visible to any user connected to these networks. Preppernau, Joan; Cox, Joyce. Windows 7 Step by Step. Redmond: Microsoft Press, Online Training Solutions, 2010, p. 85-97, 129-130.

[11] Create a Document Workspace. 2011. http://office.microsoft.com/en-us/word-help/create-a-document-workspace-HP006346865.aspx?CTT=1 (22.06.2011.). There is option to create shared workspace by using web browser and Microsoft Windows SharePoint Services Web, but only if you have permission to create Document Workspace.

Shared document spaces allow many users to work on same document and with only one (latest) version of it. There could be problem if somebody needs to work with or to view older version, because there is not useful document version history.

Groove Workspace

Microsoft Office introduces Groove 2007. Work in Groove 2007 is organized in workspaces, which provide access to information needed to be shared and tools needed for collaboration and communication during the team work. Document collaboration tools are used to share files an there are alerts when files and information have changed in a workspace and when team members perform important activities. There are tools for communication and organization like discussion tools for online conversations, such as Chat tool, meeting and calendar tools for tracking project milestones, meeting times, and updates, indicator to show online team member or contact, and there is integration with Microsoft Office Communicator 2005 and 2007 to initiate phone call or instant messaging session.[12] Groove is for smaller highly collaborative workgroup, which works simultaneously on the same file, although being made of mobile workers not always connected to the Internet. When member of Groove workspace is online he has the most up-to-date version of project documents with alerts what is changed and then he can use it when is offline. So Groove 2007 is highly effective for team members who travel and most of the time are offline.[13] Members of Groove workspace can be used as a virtual team of a people involved in a project or as members of an organization department.[14] Microsoft Office 2010 Professional Plus has SharePoint Workspace replacing formerly Groove Workspace.[15]

Groove workspace is very useful for project or team member usually mobile and working on documents using Internet connection in different time. There is possibility to have old version of his offline document if user has not been

---

[12] Groove 2007 is Microsoft Office 2007 application and as client part of Enterprise and Ultimate Edition of Microsoft Office 2007 or as stand-alone product, and as Live version. Microsoft acquired Groove Networks in 2005. Earlier versions of Groove are known as Groove Virtual Office: Standard Edition, Project Edition, and File Sharing Edition. For full functionality of Groove 2007 there are Microsoft Office Groove Server technologies. Jewell, Rick; Pierce, John, Preppernau, Barry. Microsoft Office Groove 2007 Step by Step. Redmond: Microsoft Press, 2008, p. ix-x.

[13] Jewell, Rick; Pierce, John, Preppernau, Barry. Microsoft Office Groove 2007 Step by Step. Redmond: Microsoft Press, 2008, p. 234-236.

[14] Groove Workspace basics and the Office 2007 system. 2011. http://office.microsoft.com/en-us/sharepoint-workspace-help/groove-workspace-basics-and-the-2007-office-system-HA010234124.aspx?CTT=1 (22.06.2011.).

[15] Bott, Ed; Siechert, Carl. Microsoft Office 2010 Inside Out. Redmond: MS Press, 2011, p. 8.

connected for while on Internet and on Groove workspace many changes to shared document can be made.

SharePoint Sites

SharePoint is a server-based database application that uses a Web user interface for users to access this primarily file-storage mechanism with powerful collaboration tools built in and placed on SharePoint Server. SharePoint is used for larger collaboration solutions, and in difference to Groove, has file version control with ability to check the file out and check it in. When the member of team checks the file out, it is copied to his own computer. The copy of it is locked on the server so nobody can make any changes of the file, so it becomes the "read-only" file. After the member checks the file back in, with the description of changes made, the file is available to the other team members for editing. The file version means that several iterations of a file can be saved on the server. There is an additional feature to add SharePoint Files tool to any Groove 2007 workspace with ability to publish documents to a SharePoint documents library as synchronizer.[16]

This single Web-based environment for team collaboration, communication, and information sharing easily integrates with Microsoft Office. It's based on Windows SharePoint Services, component of Windows Server 2003 and 2008, providing infrastructure for teams to create Web sites to share information and collaborate with other users. It is possible to access this stored content from a Web browser or directly through Microsoft Office. SharePoint sites provide places to share various types of information, communication, and documents, facilitate team discussions, share document collaboration, for blogging, and building knowledge bases by using wiki. Document collaboration includes checking in/out documents, document version control and recovery of previous versions, and document level security. Team members work together on documents, tasks, contacts, events, calendars, wikis, and other information. SharePoint site can have subsites hierarchical organized. Some of SharePoint site tools are shared document libraries, contacts, calendars, task lists, discussions etc. Access to these sites is controlled by a role-based system based on permissions level specifying actions which users can perform on the site: read, contribute, design, have full control, and get limited permission level.[17]

---

[16] Jewell, Rick; Pierce, John, Preppernau, Barry. Microsoft Office Groove 2007 Step by Step. Redmond: Microsoft Press, 2008, p. 234-245.

[17] Londer, Olga; English, Bill; Bleekerand, Todd; Coventry, Penelope. Microsoft Windows SharePoint Services Version 3.0 Step by Step. Redmond: Microsoft Press 2007, p. 10-17.

Microsoft Office 2007 introduces SharePoint Designer to easily manage SharePoint sites on SP Server, SharePoint Designer 2010 changes some useless features,[18] and Office 365 consists of SharePoint Online at Microsoft cloud.

SharePoint sites have history versions of documents, useful especially if somebody delete document or make unacceptable changes to older version. User cannot make changes to document if somebody else checks file out, because file is locked on server and other users can only view file.

Synchronous Online Collaboration

Microsoft Office 2010, while delivering desktop applications, introduces new concept of Office applications placed on Microsoft cloud: Web Apps.

Microsoft Office 365 is completely based on Microsoft cloud services. These online solutions can be accessed directly by using Microsoft Office 2010 applications or using Web browser. Microsoft Office 2010 has a feature of direct saving to cloud-based storage Windows Live SkyDrive on the Web.[19]

Office 365

Microsoft Office 365 consists of Internet-based Microsoft cloud services: Microsoft Office Web Apps as online companion to Microsoft Word, Excel, Access, and OneNote, Microsoft SharePoint Online, Microsoft Exchange Online, and Microsoft Lync Online. There is Office desktop setup to workstation for managing required components and updates, or can be used Microsoft Office 2010 as well as Office 2007 with SP2.[20]

Office Web Apps

Microsoft Office Web Apps is one of Microsoft cloud services available to users through Microsoft SharePoint Online and its part Microsoft Office 365, through Microsoft Windows Live, as well as through Microsoft Office 2010 Professional Plus edition (known as on-premises users).

Sharing documents is easy by using SharePoint Online document library, and easy to read and edit even with those users who don't have installed Office by using Web browser, opening and editing document directly in the browser, as well as in the Office client application.

Excel Web App has co-authoring option to simultaneously edit a workbook with people in other locations, also noticing other editor and his changes

---

[18] Coventry, Penelope. Microsoft SharePoint Designer 2010 Step by Step. Sebastopol: O'Reily Media, 2010, p. 3-8.

[19] Bott, Ed; Siechert, Carl. Microsoft Office 2010 Inside Out. Redmond: Microsoft Press, 2011, p. 299-300.

[20] Versions of Windows supported by Office 365 are: Windows XP with SP3, Windows Vista SP2, Windows 7, and Windows Server 2008. Microsoft Office Web Apps Service Description. 28.06.2011. http://www.microsoft.com/download/en/details.aspx?id=13602 (29.06.2011.).

synchronize almost in the real time. This option is available and in OneNote Web App for Shared Notebooks.[21]

Web Apps allow different users to work on the same document in the same time, with no multiplications of same document or history versions problems. There could be confusing or even frustrating if many users make changes, deleting parts of document while others work on that document. For working on Web Apps stabile broadband Internet connection is also needed.

Conclusion

Collaboration, especially the teamwork one, and implementation of projects are common in business environment nowadays the same way are the collaboration possibilities of mostly used office applications. This collaboration between different users can be asynchronous or synchronous, offline or online, client-based or server-based. Microsoft Office 2003, 2007, 2010 versions provide different application features towards collaboration.

The earliest features such as comments in Word, Excel and PowerPoint, tracking changes in Word and Excel, comparing, merging and combining documents in Word and similar ones in Excel, are used for asynchronous offline collaborations and can be found in all Office versions: 2003, 2007, and 2010.

The beginnings of collaboration using networks can be found at shared document workspaces, based on shared files, folders and files, on clients organized in the same workgroup or domain.

Microsoft Office 2007 introduces Groove 2007 and SharePoint Designer 2007, and by using Groove Workspace and SharePoint sites, it provides combination of offline and online collaboration being based on new Microsoft server-based technology known as SharePoint.

Microsoft Office 2010 Professional Plus, besides client applications, introduces cloud-based services called Web Apps such as Word Web App, Excel Web App, PowerPoint Web App, and OneNote Web App. Sharing of documents is based on SharePoint Online document library or cloud-based storage Windows Live SkyDrive. It can be used even without an Office installed, but through using web browser. Co-authoring is a feature for simultaneous editing of Excel workbook and Word document or OneNote notebook. Web Apps are available through MS Office 2010 Professional Plus, Office 365 and SharePoint Online, as well as Microsoft Windows Live.

To sum up, all Office versions provide features for offline asynchronous collaboration, Office 2007 introduces additional feature for online asynchronous collaboration, Office 2010 Professional Plus with Web Apps and Office 365 provide features for online synchronous collaboration.

---

[21] Ibidem.

References

Behera, Suresh. Office Profesional Editions VS All Other Office 2003 Editions @ VSTO. 18.04.2006. http://weblogs.asp.net/sbehera/archive/2006/04/18/443219.aspx (12.06.2011.).

Bott, Ed; Siechert, Carl. Microsoft Office 2010 Inside Out. Redmond: Microsoft Press, 2011.

Conrad, Jeff; Viescas, John L. Microsoft Access 2010 Inside Out. Sebastopol: O'Reily Media, 2010.

Coventry, Penelope. Microsoft SharePoint Designer 2010 Step by Step. Sebastopol: O'Reily Media, 2010.

Cox, Joyce; Lambert, Joan; Frye, Curtis. Microsoft Office Professional 2010. Redmond: Microsoft Press, 2011

Cox, Joyce; Preppernau, Joan. Microsoft Office PowerPoint 2007 Step by Step. Redmond: Microsoft Press, Online Training Solutions, 2007.

Cox, Joyce; Preppernau, Joan. Microsoft Office Word 2007 Korak po korak. Zagreb: Algoritam, 2008.

Frye, D. Curtis. Microsoft Excel 2007 Step by Step. Redmond: Microsoft Press, 2007.

Glenn, Walter; Northrup, Tony. Installing, Configuring, and Administering Microsoft Windows XP Professional, Second Edition. Redmond: Microsoft Press, 2005.

Jewell, Rick; Pierce, John, Preppernau, Barry. Microsoft Office Groove 2007 Step by Step. Redmond: Microsoft Press, 2008.

Londer, Olga; English, Bill; Bleekerand, Todd; Coventry, Penelope. Microsoft Windows SharePoint Services Version 3.0 Step by Step. Redmond: Microsoft Press 2007.

Microsoft Office suites. 2011. http://office.microsoft.com/en-us/suites/ (20.06.2011.).

Microsoft Office Web Apps Service Description. 28.06.2011. http://www.microsoft.com/download/en/details.aspx?id=13602 (29.06.2011.).

Microsoft Windows XP Korak po korak. 3rd ed. Zagreb: Algoritam, 2005.

Murray, Katherine. Microsoft Word 2010 Inside Out. Sebastopol: O'Reily Media, 2010.

Office Professional Academic 2010., 2011. http://office.microsoft.com/en-us/professional-academic/ (20.06.2011.).

Preppernau, Joan; Cox, Joyce. Windows 7 Step by Step. Redmond: Microsoft Press, Online Training Solutions, 2010.

Preppernau, Joan; Cox, Joyce. Windows Vista Step by Step. Redmond: Microsoft Press, Online Training Solutions, 2007.

Thurrott, Paul. Microsoft Office 2003 Editions Compared. 06.10.2010. http://www.winsupersite.com/article/product-review/microsoft-office-2003-editions-compared (12.06.2011.).

Which suite is right for you? 2011. http://office.microsoft.com/en-us/buy/office-2010-which-suite-is-right-for-you-FX101825640.aspx (20.06.2011.).

E-LEARNING

Between Cooperation and Conflict in Quality Assurance: Principles of Toyota's Just-In-Time Production for Training Geriatric Staff

Max Liebscht

LISL - Lausitzer Institut für Systemischere Lösungen
Rosenstrasse 1, 02826 Görlitz, Germany
Mail@angewandte-psychologie.com

Martin Weitzmann

LISL - Lausitzer Institut für Systemischere Lösungen
Pablo-Picasso-Strasse 1, 13057 Berlin, Germany
Mail@welearning.eu

Katrin Schubert
Fernuniversität Hagen
Hauptstrasse 163, 02788 Zittau, Germany
Mail@welearning.eu

Summary

Due to rising cost in geriatric care, discussing, defining and updating consistent quality standards is of utmost importance. E-learning in the cloud can provide synergies between all stakeholders in quality assurance so that the transfer of theoretic knowledge into practice will benefit in ways which will sustainably enhance care quality for the elderly. Consistent use of the findings from the stakeholder analysis results in a transfer of technologies and communication architectures from the traditional business sector to the non-profit sector. The discussion will focus on two aspects transferred from Toyota's just-in-time principle: focusing all parties involved by decentralized provision of relevant information and discourse loops between prosumers for virtual assistance.

Key words: e-Learning, e-services, business applications, principle of just-in-time-production, prosumers

Introduction
Knowledge management in learning organisations (Senge, 2004) and stakeholder networks can be substantially modified by information processing systems (Kruse, 2004). This understanding and our background as trainers, systemic organisation developers and future care patients raises the following question: How can we develop a virtually assisted communication design

189

facilitating the integration of care quality for all parties involved in geriatric care? If geriatrics training and studies can be brought up to date in a consensus between all stakeholders, new standards can be transferred to practical use. To optimize any form of communication it is first necessary to analyse the parties involved. Which interest groups and institutions are involved in geriatric care? Which are the most essential problems for each of these groups? Once these problems are identified, they can be used as links to create optimized communication processes. In the following, nineteen parties are identified which are players in geriatric care and whose least common denominator is the interest in defining what up-to-date quality in geriatric care should, could and must achieve.

Decentralised provision of relevant information as a focus for all parties involved

a.) In theory, people in need of care are the focus of all caring activities, i.e. stakeholders are to concentrate their work on this group of people which is in particular need of protection. Due to their health and legal situation, they are often hardly able to take care of their needs.

b.) Family members, who are often quite unexpectedly faced with care issues, seem to require a large amount of orientation. The largest challenge is to find the information best suitable to the situation and personal needs in the vast amount of data provided. Furthermore it would be interesting for them to know what professional care workers learn during their training in order to assess the know-how to be expected. The transparency of training courses, however, is hardly given.

c.) Care providers usually act within the legal framework provided by political representatives perhaps competent in care issues. Unfortunately, the trend is to transfer business with obvious profit potential to private-sector organisations. Classic non-profit tasks are more and more transferred to profit-oriented service providers although these tasks have been efficiently and sustainably been solved by families and later on by church-run organisations. Losses are borne by the public and profits are privatised. Due to the demographic development the profit expectations involved in geriatric care are now fulfilled often only at the expense of defenceless people. News headlines cannot substitute statistic analyses; they can, however, indicate the temptation (DRK, 2011) managers of care-providing facilities might yield to. One basic temptation could be found in reinvesting earnings to improve balance sheets instead of upgrading quality of the services provided. Even major geriatric care providers are again and again discovered to use earnings for share speculations (Pubantz, 2009) and not to stock up reserves for hard times. Decreasing fees for geriatric institutions are never heard of although well-planned investment strategies and budgeting might enable such a scenario. Charitable organisations tend to benefit from subsidies, the profits earned, however, are rarely transferred back to the public even if this

190

might be expected from a macro economic perspective. Discussing quality issues more in public may create here a higher degree of sensitivity (EpD, 2010) in decision making and improved social control (Birgelen, 2009) by the general public.

d.) The ways of assuming managerial responsibility suggested above provide the framework for executives in care institutions when they find compromises between so-called practical constraints and ethical requirements in care quality on an every-day basis. It is the issue of labour costs or client-staff ratio which is again and again under discussion in geriatric facilities.

e.) Committed staff in training geriatric homes faces the problem of their career advancement on the one side and is often left alone when implementing professional standards or they are in need of substantiating their approach. Who the professional authority encouraging them?

f.) Unskilled geriatric care assistants are on the lowest level of hierarchy seen from a formal point of view. "Proletarising care services" is on the rise in spite of all reservations. It seems to be inadequate that assistants are not only badly paid but also badly trained. These assistants take care of a major part of the practical work in care facilities, but have nearly no lobby at all except with the people they take care of.

g.) State-run organisations in health care politics deal with care provision or education in geriatric and care issues, as well as with care quality in geriatric homes and often mediate between the different interests. The quantitative assessment scheme for geriatric homes carried out by the Medical Review Board of the Statutory Health Insurance Funds in Germany introduced in Germany in 2010 (Medizinischer Dienst der Krankenversicherung. Qualitäts-prüfung von Pflegeeinrichtungen, 2011) was widely criticized on the background that homes which were regionally known as practising "dangerous care" were assessed as being "very good" homes after arranged quality checks (BKK Bundesverband GbR, 2011).

h.) A further target group for quality improvement and qualification are university students in nursing studies. This field of studies is mainly offered as an extra qualification and its quality essentially depends on the degree academic approaches taught reflect every-day life in care facilities with its ups and downs.

i.) Which is the particular interest when classifying graduates as a target group? In theory one might expect them to be an especially welcome asset on any ward as it is them who bring the up-to-date knowledge in care and professional standards to the wards. From a social psychological point of view the enthusiasm of long-served staff is quite limited. This is not only due to the fact that potential future managers are first informally tested on their loyalty. New team members rank low in any informal hierarchy. As the relationship aspect in interpersonal communication is more important than the factual aspect, new input of graduates is appreciated as an exception only. If transferring theoretical

knowledge into practical use is not to result in excessive stress for the individual, those ranking lower in status need authorities and an institutional framework during this transient period in order to have their commitment supported.

j.) Conflicting requirements in care institutions are reflected in training organisations like a mirror. Which are the legal and communicative frameworks necessary for education service providers so that they can convince markets with setting a good example in quality standards and well integrated curricular offers and perceive pedagogic and economic aspects as of equal importance?

k.) Which form of help is essential for managers in training organisations when they recognized moderating knowledge management for their staff as part of their leadership function?

l.) Support is also necessary for permanently employed staff if they would like to keep up their input in the courses they teach and to combine their expert authority with course contents which is transparent for all involved, i.e. teachers AS WELL AS trainees.

m.) Freelance teachers who teach psychological, sociological and communi-cation subjects in theory and care practice have normally not a profound knowledge in the field. It is important to note that permanently employed teachers and freelancers see themselves usually as competitors a fact which complicates transparent knowledge management and integration enormously. When asked to share their know-how, they will be afraid of being replaceable.

n.) The most profound problem students in vocational schools face is the fact that they are expected to be fascinated by questions which they have usually not encountered yet in their everyday life. Teachers provide them with answers to decision-making problems they have never dealt with as they have not had hands-on experience in care homes, especially in their first year of training.

o.) Private-company staff in charge of certifying quality standards are usually not sufficiently acquainted with the field they certify in so that they are not able to assess the quality of cooperation between teachers with different employment status, i.e. to what extent the courses are integrated instead of fitting just into different spheres of knowledge. However, more than ninety different certifying products (Medizinauskunft, 2004) are offered in Germany in the geriatric care sector by legally authorized companies. There seems to be an urgent need to have certifying agencies evaluate more competently to what degree the courses offered by training providers in care training are up to date according to latest findings in care sciences. From an official point of view this necessity does not seem to exist as it is unrealistic to update the legally binding curricula involving all the authorities needed to bring such changes under way. Could open access content and web-based discourse systems provide fresh impetus to further quality standards?

p.) University staff doing research in geriatrics and care faces the challenge to leave their ivory tower and to do participating field research. Such type of

research needs exchange with staff working in geriatric care so that new approaches create more than a storm in a teacup for care organisations.

q.) Editors and other commercial forms of knowledge providers base their sales strategies on the need for orientation of all the different stakeholders. Charitable organisations and foundations, e.g. the German Board of Geriatric Care approach profit orientation in more relative terms.

r.) The media more often takes up subjects of quality in old age and care for their main target groups due to recent demographic trends.

s.) Last but not least we classify so-called public top performers who are getting used to issues of getting older and being in need of care despite the trend to block these subjects out. Qualified new performers can be found only in their ranks, bringing the suitable maturity for geriatric care – a profession still having a bad public image.

The fragmentations most essentially found in communication could not be dealt here in greater detail. The main problems, however, can be summarized as follows: On the one hand, there is no medium where all the stakeholders can get to the point in discussing quality standards. On the other side, interests and values seem to be highly contradictory for the institutions and stakeholders involved in quality management. Training providers run by private organisations or local authorities see each other often as competitors for recruiting students. The present culture of communication between permanently employed and freelance teachers results in the fact that students feel that the courses on offer to be fragmented and not integrated. All the individual groups listed above have their least common denominator in the fact that they all will find themselves in the middle of fig. 1. Does this fact already show the way to solve the problem? The circle shown in Figure 1 represents a (incomplete) macro-system of all social interest groups involved in the "problem" of care quality with the point of focus right in the centre.

The diversity of target and interest groups involved in an intervention into this social area of conflict reflects the social and political complexity. How can such a complexity be reduced without being trivialised? The first aspect of our approach to this question is to establish a common focus of communication between the different interest groups. The second aspect can be derived from Toyota's just-in-time principle introduced in the 1950s. If care institutions orient themselves more towards market-oriented rational criteria than to charitable views, it only seems to be consistent to make the best of it and to learn from efficient or sustainable organisational principles of businesses.

The just-in-time principle aims at meeting the demand exactly when needed in the quality and quantity required at the point needed. The Japanese production and logistics strategy is to consider the entire order process in a holistic way. With knowledge products mainly based on information, this process has been dramatically simplified thanks to the Internet synchronising the individual

functions within the value chain up to the finished product. Materials flows found in the traditional publishing industry no longer play an essential role. The dominating role of traditional elite universities which have achieved their success in a grown infrastructure with the help of business partners can be set off by "universities in the cloud" which provide excellent expert knowledge and didactics despite their physical location far away from urban centres. The missing link has just been the consistent transfer of JIT's success (together with successful principles developed on this bases such as module-specific feedback loops in KANBAN) to a web-based software architecture synchronising experts defining demand and experts developing products and manufacturing them. As a consequence quality assurance should be handed over to a moderated discourse between former producers and consumers, as these parties can find maximum agreement for their requirements. The coordination between contents authors and licensees will be moderated by editors, who are also in charge of maintaining the communication design. The main focus is put on scientific state of the art, usability and efficient didactics.
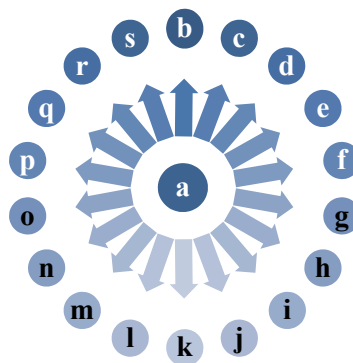
Fig. 1: System (Schmidt, 2005) of social groups involved in defining quality standards

When qualifying care staff analogies to JIT principles can be found: all stakeholders focus their commitment on care quality which is in news headlines again and again as quality is demanded by the public. When bringing theory and practise of qualifying executives and professionals together, the point in question is to bring apparent opponents, i.e. manufacturers and consumers, together and integrate all stakeholders as prosumers (cf. Toffler). Prosumers act together in defining their current requirements as exactly as possible and in co-producing the final product. Cultivating the dialogues between suppliers and manufacturers and subsequently producers and consumers will make warehousing costs redundant and avoid on-spec production. Nowadays it is essential to avoid storing knowledge which is no longer up to date in its

194

expertise and does not meet the requirements of practical decision making. The product in focus is up-to-date knowledge prepared with logically consistent didactics inviting the user to participate in learning and teaching. The discourse system we are developing is virtually assisted and can be used beyond Saxony to transform the stakeholders mentioned above into prosumers. More details of the project and its architecture are available in Liebscht, Weitzmann & Schubert, 2011.

Modes to virtually assist discoursivity

Buber's dialogue ethics, synergetics as well as Toffler's concept of prosumers in discourse inspired the orientation towards calibrated feedback loops.
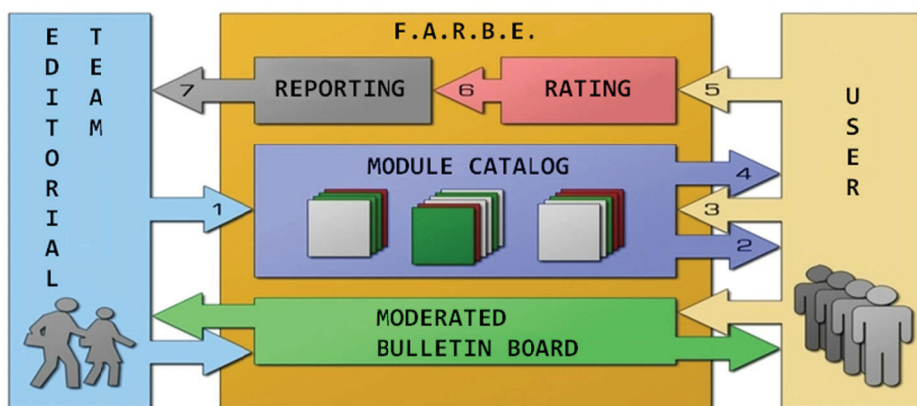
Fig. 2: Process chain for a geriatric knowledge source

Process phase 1: A team of experts in geriatrics creates an initial version of learning modules according to the didactic F.A.R.B.E principle (Liebscht, Schubert, Weitzmann, 2011) together with graphic designers, programmers, photographers and video producers. This initial version is commented on module by module by experts. In contrast to knowledge sources such as Wikipedia, the learning content is consistently organised in a) typical decision-making situations in geriatric jobs, b) embedded into model social contexts using role plays and c) on a firm scientific basis. The decision-making situations are prototypes and can be varied according to practical contexts. Experts, wishing to sharpen their profile in their field and in didactics, can use the prototypical modules to create their own content, have it evaluated and optimized thanks to the feedback loop.

User feedback is the cornerstone of quality enhancement and keeping topics up to date. At the beginning of each training session, a clearly structured entry mask has to be filled in. To keep on evaluating the system it is essential to learn

the target group of the user keying in the feedback. To give an example: What is the his/her practical experience when rating practical relevance? When interpreting the qualitative assessment data, the editorial staff can take into consideration whether a care manager, a family member with someone in need of care, a care assistant or a freelance teacher evaluated the quality of content and didactics of a module. If a user works on the modules from the same computer, user identification can be stored and the input mask skipped.

Process Phase 2: Users having registered use the modules and fill in the test. Modules can be selected at random, in chronological order or according to their level of difficulty. A further option is to deal with the modules according to the German-wide curricula for geriatrics care (Kuratorium Deutsche Altershilfe, 2002). Technical terms can be found through a search function. Modules the user has not dealt with to his/her satisfaction can be repeated. In this sense, the system can be used according to the flashcard principle. This is a systematic principle based on progress, optimizing learning in each successive learning session: all modules where the incorrect test answers were given have to be repeated until all test answers are correct.

Process phase 3: The modules can be used as an interactive textbook, as a reference or a model for role plays. In figure 2 processes indicate the use as a test forum. Learning by heart without understanding the context is not given as a large number of questions are randomly generated and always embedded into job-related contexts.

Process phase 4: The test scores of the modules (sorted into correct and incorrect answers) enable younger users in particular to realistically assess their current level of geriatric competence in good time and in relation to practical relevance. The answers given show the users in need to pass an examination for a geriatric qualification what he/she needs to learn in order to pass the test.

Process phase 5: The five components of a module to be dealt with online correspond to one screen mask each. To proceed to the next component, the user needs a mouse click. This click will be used to evaluate the module component on a four-item rating scale, with one corresponding to "very good", 2 for "good", 3 for "unsatisfactory" and 4 for "fail". See the following five rating questions as an example:

F: To what extent is the situational context for the question concisely described?

A: To what extent do the possible answers reflect decision options in practice?

R: To what extent can you comprehend the views and conflicts of the people described?

B: To what extent do the examples of job-related topics correspond to the question?

E: to what extent do the substantiations for correct and incorrect answers the state of the art in geriatric care?

Process phase 6: The quantitative rating given in an input mask can be complemented by written statements. Such qualitative comments substantiating the quantitative rating score double when interpreting the rating.

Process phase 7: The reporting system described provides clues for the author, editing staff und users to see which modules are in need of optimisation. The urgency of updating a module is shown on a scale working like a thermometer. The qualitative comments given by prosumers indicate the details to be optimised. Working like an online exchange, the weekly price for the modules is linked to the quality rating given by previous users. Authors with a reputation for being committed to optimize their modules will gain from their expertise and didactic input by increased demand for their modules and higher prices. Training organisations which buy licences for module sets early on can be sure to have made a safe investment as the modules are continuously optimized and updated independent of the price they have paid once for the modules. This system fosters and promotes high-quality didactics, which is rare to find as a high amount of input is required.

Conclusion

When transferring the basic principles of Toyota's just-in-time production to distribute and update knowledge relevant to geriatric practice, discussing quality standards and quality assurance go hand in hand. Each interest group may save their face and no stakeholder participating in the discourse needs to be afraid of providing advantages to competitors. Everybody involved can learn from each other and together update quality standards in geriatric care thanks to the online communication design introduced in this paper. The system contributes to better combine economic and psychological rational criteria for quality development. As this communication design sustainably assists discoursivity, significant improvement in geriatric care is to be expected thanks to care staff being well informed. The pilot project aims at transferring the principles applied to other fields such as educating educators and precise definition of requirements for junior management, among others. Combining nearly all corporate functions of the process chain will be transferred to the field of knowledge reproduction in the publishing sector due to advanced technology in the near future. Further modes of the discourse systems and technological challenges are described in a book accompanying the project. A beta version with a sample of 100 modules is expected in March 2012.

References

Birgelen, Dominik. Volkssolidarität Fonds insolvent. 2009. http://www.senioren-blogger.de/recht-verbraucher/volkssolidaritaet-fonds-insolvent.html (19.07.2011)

BKK Bundesverband GbR. Qualität der Pflegeeinrichtung. 2011. http://www.bkk-pflegefinder.de/details/index.php?uid=511420146_02791_ams_s (15.07.2011)

Bosch, Bernhard. Handbuch Projektmanagement für ESF-Projektträger. Stuttgart: Deutscher Paritätischer Wohlfahrtsverband, Landesverband Baden-Württemberg e.V., 2007.

Buber, Martin. Das dialogische Prinzip: Ich und Du. Zwiesprache. Die Frage an den Einzelnen. Elemente des Zwischenmenschlichen. Zur Geschichte des dialogischen Prinzips. Gütersloh: Gütersloher Verlagshaus, 2001.

Deutsches Rotes Kreuz. 23. April 2011 - Neue Wege vier Jahre nach Skandal. 2011. http://drk.ac/aktuelles/newsdetails/archiv/2011/april/23/meldung/60-23-april-2011-neue-wege-vier-jahre-nach-skandal.html (18.07.2011)

Evangelischer Pressedienst. Ex-Diakoniechef hat weitere Spendengelder unterschlagen. 2010. http://www.diakonie-oldenburger-land.de/index.asp?ID=17697 (18.07.2011)

Haken, Hermann; Schiepek, Günter. Synergetik in der Psychologie. Selbstorganisation verstehen und gestalten. Göttingen: Hogrefe. 2010.

Kruse, Peter. Next Practice. Erfolgreiches Management von Instabilität. Veränderung durch Vernetzung. Offenbach: GABAL, 2004.

Kuratorium Deutsche Altershilfe. Bundeseinheitliche Altenpflegeausbildung. Materialien für die Umsetzung der Stundentafel. 2002. http://www.kda.de/files/curriculum/curric-kda.pdf (19.07.2011)

Liebscht, Max; Schubert, Katrin; Weitzmann, Martin. The F.A.R.B.E. eLearning system. 2011. http://www.pixel-online.net/edu_future/common/download/Paper_pdf/ENT23-Liebscht.pdf (14. 07.2011)

Liebscht, Max; Weitzmann, Martin; Schubert, Katrin. Gemeinschaftsprodukt Pflegequalität - Das Kompetenz-Training-Pflege als Modellprojekt zur integrierten Kompetenzentwicklung von Fach- und Führungskräften am Beispiel Altenpflege. 2011. www.systemagazin.de (in prep.)

Medizinauskunft. Gütesiegel in der Pflege: Keine Garantie. 2004. http://www.medizinauskunft.de/artikel/special/19_08_pruefsiegel.php (18.07.2011)

Medizinischer Dienst der Krankenversicherung. Qualitätsprüfung von Pflegeeinrichtungen. 2011. http://www.mdk.de/1328.htm (15.07.2011)

Pubantz, Frank. Der VS-Skandal. 1386 Anleger geprellt. // *Der Priegnitzer. (*Ausgabe 1.9.2009), http://www.prignitzer.de/nachrichten/lokales/guestrow/artikeldetails/article//vs-skandal-1386-anleger-geprellt.html (18.07.2011)

Senge, Peter M.; Kleiner, Art; Roberts, Charlotte. Das Fieldbook zur 'Fünften Disziplin'. Stuttgart: Klett-Cotta, 2004.

Schmidt, Gunther. Systemische und hypnotherapeutische Konzepte für Organisationsberatung, Coaching & Persönlichkeitsentwicklung. (8 MP3-CDs) Müllheim- Baden: Auditorium-Netzwerk, 2005.

Toffler, Alvin. Die dritte Welle. Zukunftschance. Perspektiven für die Gesellschaft des 21. Jahrhunderts. München: Goldmann, 1988.

WIdO (Hrsg.); Gerste/Schwinger/Rehbein. Qualitätssiegel und Zertifikate für Pflegeeinrichtungen - Ein Marktüberblick, Bonn WIdO-Materialien Bd. 51, 2004.

iPad: A New Classroom Technology?
A Report From Two Pilot Studies

Alma L. Culén
Department of Informatics, University of Oslo
P. Boks 1080, 0316 Blindern, Oslo, Norway
almira@ifi.uio.no

Andrea Gasparini
Digital Services, University of Oslo Library
P. Boks 1085, 0317 Blindern, Oslo, Norway
a.a.gasparini@ub.uio.no

Summary

In this paper we discuss two pilot studies involving the use of iPads for active reading in a teaching/learning situation. This is part of a broader study of how introducing tablet PCs may transform the work and learning practices of learners. One of the pilot studies was conducted in a graduate level course, involving 40 university students. The other study involved 26 fourth grade elementary schoolchildren. The results concerning acceptance of the technology were vastly different in the two studies. We find the comparison to be very interesting in several aspects, most notably on the issue of ownership and perceived usefulness. We hope that our experience with these pilot studies may be of use and interest for a wider community. Our research method is based on ethnography (in-class observations), enriched by workshops, questionnaires, group and individual interviews involving students, faculty and, in the case of elementary schoolchildren, families. The data from interviews has been consolidated and mapped out into an affinity diagram. The resulting diagram shows clearly issues that should be further addressed, as well as areas where changes in study-related work practices may occur. This paper offers some reflections on differences and similarities observed in the two study situations.

Key words: Classroom Ecology; iPad; Learning; Mobile Technology in Classroom; Tablet PC; Technology Adoption; Touch User Interface.

Introduction

A system consisting of students, teacher, practices, values and technology may be referred to as classroom information ecology (Nardi, 1999). Our interest was to observe how introducing a tablet, in this case an iPad, would change that ecology. Unlike more traditional 'desktop technologies', mobile technology like

tablet PCs may be easier integrated into the daily life of students. It also has the potential to redefine what constitutes a learning space. Without constraints of specific time and place, it may facilitate more robustly situated learning practices. Whether this hypothesis holds true needs to be verified in real life situations. Many educators world around are aware of this potential and there are many studies that are being conducted around the globe testing this or a similar hypothesis (see (Hu, 2011), (Chen, 2010), (Wilson, 2011) or (White, 2010)). In (Vollen, 2011), Danish IT paper, it is reported that the rector of the school conducting one such experiment in Denmark has said: "It does not sound nice, but we'll see if we can claim a larger portion of students' free time. The path to learning is now shorter. Students may, whenever the opportunity arises, read school related texts or watch a videotaped lectures in the comfort of their sofa in the evening." (Trans. Culén). However, scientific studies on the effect of iPad-based learning are yet to come.

An opportunity to conduct the two pilot studies in a real learning situation arose in the fall semester of 2010, when the University of Oslo Library decided to try out a digital curriculum on iPad, and equipped an entire Geology class (40 students and their instructor) with iPads. In the spring semester of 2011, some of these iPads were used in an elementary school study. The goal of the studies was to see how students adopt this new technology and how does it influence the classroom ecology. The first pilot study was also part of the (Green University, 2010) project with focus on the environment. The use of paper and the volume of printing was the primary measure of possible environmental benefits of using iPads in the class.

The Geology class was chosen without any special considerations as to how iPad could be used in that field. However, placing an entire course curriculum on the iPad at no expense for students, alongside of environmental benefits of less printing, was viewed as something that would increase the perceived usefulness of the tablet for students. In addition, iPads are generally viewed as devices that have an easy to use, intuitive interface. In (Davis, 1987), the influence of perceived usefulness and perceived ease of use are discussed in relation to acceptance of technology. Many other variables have been relevant to the technology acceptance model (TAM). In (Barki et al., 2007) the authors name: trust, cognitive absorption, self-efficacy, job relevance, image, result demonstrability, disconfirmation, information satisfaction, top management commitment, personal innovativeness, information quality, system quality, computer anxiety, computer playfulness, and perceptions of external control as some of the factors that may be important for the acceptance of technology. Our findings show that while many of the above showed up in the course of the study, ownership and the possibility of sharing the work on iPad have been the most important (Culén et al., 2011). However, at the end of the pilot study we could only conclude that the introduction of this technology had not been a success with the Geology class. The disinterest of students in this platform for work purposes is illustrated

by the fact that all of the students were offered the opportunity to buy the iPads that they used throughout the semester at a favourable price, but only 3 students took up the offer.

On the other hand, the results of the second pilot study show that the iPad has been successfully introduced as a tool at an elementary school. As an indication of how important the iPads were for children, we can quote one of the fourth graders who participated in the study: "*the best things we ever had at school are iPads and chickens*" (they have incubated some eggs and watched the chickens get out of their shells in the second grade).

Context of Pilot Study 1

Forty students, 1 lecturer and 3 teaching assistants participated in the study in the fall of 2010. Each iPad had the class curriculum downloaded in advance. The curriculum for the course consisted of book chapters, lecture slides, maps and academic papers. Each iPad came with a Dropbox containing the curriculum. The students have also received a gift card of approximately $25, and were required to get iAnnotate and Elements applications from the Apple Apps store which would enable them to add their own annotations, highlight the text etc. No stylus or cover for multi-positional viewing of iPad was given to the students. The physical setting for this course was typical for higher educational institutions countrywide. Students have lectures in a large auditorium and discussion/work groups in small groups and rooms. Wi-Fi is available everywhere at the University premises, but student housing, where many international students from this class lived, did not have Wi-Fi, thus disabling iPad Internet use while studying at home. Their program is very competitive and fast paced, thus leaving students with little time for anything else but studies. The students have signed an agreement to participate in our study, and committed to participation in one workshop and two surveys.

Context of Pilot Study 2

Six iPads were given to a class of 26 children: 1 for the teacher and 5 for the children. The study began in January 2011 and will last for one year. At the end of the first semester, we could report that the iPads have been successfully integrated into the classroom ecology (Gasparini et al., 2011). The schoolchildren use a spacious classroom, equipped with a Smart Board, laptop (usually connected to the Smart Board and used exclusively by the teacher) and three stationary PC's for student use. This is standard equipment for classrooms at this school and a common setting for other elementary schools countrywide. There was no Wi-Fi connection in the classroom; wireless mobile broadband was installed for the purposes of this study. As the iPads were also to be taken home, it is relevant to note that all participants had a wireless network at home and access to either a PC or a laptop. The children are 4th graders (aged 8-9). They are technologically savvy (see (Buckingham, 2007) and (Druin, 2009) on facets of

children's involvement with technology). The authors of this paper had used this same class in a previous study (Culén et al., 2011) involving the children in the co-design of an e-book reader interface. A digitized curriculum is not yet common in elementary schools. However, access to a digitized curriculum was obtained from the academic publisher (free of charge) for Religious Studies, Mathematics and Science. English is relevant both as a school subject and as the language of the applications. The children have some knowledge of the language, but many are far from fluent. The traditional way of teaching English was supplemented from the start of the study with stories and Apps that could help children to improve their English through play. No restrictions were imposed on what they could download and how they could use the tablets in their free time. Each iPad had an iTunes account with $25, with no required purchases. The children were left to make a financial decision: if they wanted more expensive Apps, they could join forces and pay for them as a group, or find other ways of managing the funds. Each iPad came with a Dropbox containing the curriculum books and the iAnnotate application, which enables users to highlight the text, make notes, etc. The children, like Geology students, were not given a stylus or a cover for multi-positional viewing (see Figure 1). The parents of the schoolchildren have signed an agreement giving us permission to conduct surveys, workshops and short interviews with children at school. Two families have agreed to be interviewed in their homes.

Figure 1. Students received an iPad each, the schoolchidren one iPad per group of five children.

The Method
The main method was based on ethnography. We were aware of technology adoption assessment questions such as those in (Staley, 2004). In Pilot Study 1, graduate students of informatics worked with students of geology during the semester. The students of informatics observed the use of iPads in the classroom, carrying out a contextual inquiry, and also doubling as technical support.

Additional data was collected from two surveys, one workshop and three group interviews (2 interviewers per group and 4-6 participants). After the end of the course, 3 students and the instructor were interviewed individually. All interviews, group and individual, were recorded and transcribed. The interview data was consolidated using the sticky notes method (one observation per note) to map the observations into an affinity diagram. The analysis pointed towards ownership and cooperation as new and interesting variables to consider in relation to mobile technology adoption in education.

In Pilot Study 2, for the first weeks of the study, we followed the grounded theory procedures (Sharp et al., 2007; 388) and simply observed the ecosystem, waiting for participants' main concerns, challenges or areas of mastery of something new to emerge. The children were observed in the class every Tuesday. The researcher also provided technical support. In addition, we conducted two workshops, interviews with two families (including students), as well as one with the teacher. Data was collected using audio, video, notes, photographs and periodic collection of iPads in order to view and document the content. Many informal conversations during the observations were very valuable. Some quantitative data was collected through four short surveys (each comprising 1-5 questions). The data was analyzed and categorized similarly as in Pilot Study 1. The analysis indicated three variables of particular interest: use of the iPads for creative learning, attitudes towards learning, and the emergence of new social patterns.

Organizational challenges

This set of challenges addresses issues around the premises on which the iPads are distributed (short term loan, long term loan, owning) to students, how the content is acquired and later accessed, who is to provide the support, and when.

In both studies, participants are "borrowing" iPads for a given length of time (one semester for students and a whole year for schoolchildren). In both cases, technical support was made available to all participants. Both students and schoolchildren needed support of various kinds, most notably with equipment breakdowns (one iPad stopped working completely, but many participants experienced temporary problems when iPads were not shut down for a long time). The students needed a tutorial on iAnnotate (YouTube, 2010); the elementary schoolchildren were given one hour of introduction to iPad at the beginning of the semester, and once a week they could get help with whatever problems they had, most often with the wireless network connection and with download of Apps as well as some guidance on how to use them.

The most important variables that were directly, but only partially, related to organizational challenges were:

- Perceived intuitiveness and the ease of use of iPad
- Perceived ownership

General perception that touch user interfaces (TUIs) are natural, intuitive etc., was falsely extended into thinking that applications would be equally easy to master by graduate students. However, they needed guidance on the use of the basics they got with the iPads, such as Dropbox and iAnnotate. The students felt that learning all these apps "properly" would take too much time. Therefore, in order to make them more willing to set the time they need to get used to working with iPads, a tutorial for iAnnotate was made (YouTube, 2010). All of this was perceived as rather complicated to use, in spite of the fact that the TUI itself was found to be very easy by majority of students.

The children, on the other hand, were not under time or academic pressure. They were interested in exploring and found it not to be difficult at all. Thus, for the level of tasks they were performing, they found the iPad to be easy to use, intuitive and playful. It is perhaps interesting to remark that arrangements around sharing of the 5 iPads among the children have gone without any problems and were fully self regulated. They have never complained about someone doing something on the iPad they did not like (such as removing content they placed on it).

When it comes to ownership, a more detailed report may be found in (Culén et al., 2011). It suffice to say that the schoolchildren live "in the now", and the timeframe for which they could use the iPads did not weigh on them. Neither were they concerned with the destiny of their work stored on iPads. They are happily unaware of many aspects of the ownership issues, and this variable was of no importance for them. Quite a different picture is seen when it comes to students. Again, part of the problem could be resolved by organizing the terms of the loan of iPads to students in a different manner (as was done at Stanford University (Hussein, 2010)), but other ownership issues, such as proprietary (Apple) software or ownership of annotations made on PDF files and stored in the cloud, would still persist.

Challenges due to physical environment

In this category, findings were also quite different. Students have quickly found out that taking notes on iPads is hard, not only because iAnnotate is difficult to master, but also, because their physical space is limited to a chair with a small work surface, which is insufficient for holding an iPad, a book and some paper and pencils. They were much happier with use of iPads in smaller discussion rooms with tables, where they could share images and texts from their iPads. Apart from this, as aforementioned, the availability of Wi-Fi on iPads was limited to the University, thus forcing the students to use devices that could be connected to the local area network by wire.

When it comes to elementary schoolchildren, the adjustments and changes in the physical environment they had to make due to the number of iPads they got (the classroom was now organised into 5 large work areas, one for each group of children with an iPad they could jointly use) fostered collaboration and

sharing, and increased the interactions among the children. They had also the only Wi-Fi equipped classroom in the school (enabling increased use of the internet in class). The Wi-Fi did not work perfectly, but everyone was very patient with it, indicating that the benefits outweighed the problems.

Academic challenges

When a new device is put into a classroom use, it naturally changes the way students work. In Pilot Study 1, we found that time pressure and the need to obtain a good grade in the class were factors that prevented students from making much time to explore the possibilities that iPad offers. Their field has a strong tradition of how the things are done. Students often resorted to these traditional means (please see the Survey2_Geology, 2010) thus missing the benefits of some features iPads offer. For example, none of the university students searched the curriculum on their iPads for specific themes or concepts, or shared their own notes taken during the lectures via email or Dropbox.

In Pilot Study 2, the challenges concerned the selection of appropriate educational applications that could adequately supplement the teaching, a common theme for many of the studies concerning the iPad in education.

It is well known that the role of the teacher in acceptance of new classroom technology (see for example (Baylor et al., 2002)) is very important. The teachers in the two studies were providing different role models for their respective students:

- The University professor has a well-established course, with a long tradition, and learning to use an iPad efficiently would take a lot of time: "*When I have very long working days and I want to be as effective as possible, the effort of sitting down for 2 -3 hours to learn the iPad is too great for me*". (Trans. Culén). He did not use the iPad when teaching.
- The elementary school teacher used the iPad actively every day during classes for variety of tasks.

Technological challenges

These challenges were of much greater importance to students and they have given a long list of frustrations, some of which are:

- Two applications cannot be open at the same time (for example, it is not possible to follow the lecture slides and browse at the same time). Students are used to multitasking in this field.
- Reloading pages or slides in PDF format takes a very long time (for example, if the text references some figure that is on a different page, it can take a considerable time to find the figure; similarly when zooming on a figure, which Geology students often do, it may be slow to reload).
- Downloading files was difficult for many students.

205

- No support for Flash.
- Problems with 3D viewing.

Some of the challenges that emerged during the course of the two pilot projects were not too "serious". They could be eliminated or bypassed. Others could be resolved when iPad is redesigned, or when some other tablet offers better support for active reading.

Samples from interviews and surveys

Surveys were focused on both environmental questions (as mentioned in the introduction, mainly related to the reduction in printing) and on use for course work. Regarding the use of paper, we see a positive trend in both studies. The students have achieved a significant reduction in the amount of pages they printed for the Geology class. 57% of respondents have answered that they have printed much less than usual. Before the iPads, the school children used to make copies of the book pages in order to do exercises, but with iPads, they had no need to make extra copies. Thus, they too, have reduced printing significantly.

The first survey at the elementary school was held after the participants had been using iPads in class for one month. Just before the survey, the children had a read-aloud session. Each child had to read a particular story from the iPad for the rest of the group. After the reading, the children discussed the story. The survey was based on simple questions that were to be answered with a star rating (5 stars being the highest score and one star the lowest). Twenty children were at school that day, and 17 rated the iPad as a preferred or equal platform for reading. Only three were negative toward the reading experience on this new artefact. After four months, we asked if they had changed their minds over reading preferences. Twenty students were present: two had changed their mind in favour of books, while two had changed in favour of iPads. The balance remained the same: 17 for iPads, 3 for books. One of the participants who chose books did so on the basis of a usability issue with iPads: it was too easy to change the page and the student kept doing so accidentally. One of the students who chose the iPad on this occasion had voted against it previously.

Through the interview with the elementary school teacher we find that she considers iPads as both useful and enjoyable to have in the classroom, even though in some cases the results were poorer then when using traditional methods. An example here was an attempt to teach them about composition with the help of iPad App Puppet Palls designed for storytelling. While they were very engaged and entertained, the teacher evaluated the results as weaker in terms of learning outcome.

When asked if preparation for the class is more difficult now that she also needs to think about iPads, she said: "*No! It in fact simplifies matters. I can ask them to use their iPads to check things and they should manage to do it by themselves. They have tools for doing it by themselves (referring to Apps, Wikipedia or Google search). ... They are also better at reading from the screen (referring*

to iPad). I skim-read, but they certainly get interested and immersed, and they actually got things from one text we were reading that I did not get, as I did not read carefully enough." (Trans. Culén).

When it comes to reading, 85% of children answered that they prefer to read from the iPad than from the book. They liked the ability to zoom in and out, and while reading, many of them were changing scopes often.

The data on the reading habits of the university students can be found in surveys. The two surveys are provided in their entirety at (Survey_Geology, 2010) for the first one and (Survey2_Geology, 2010) for the second one. Some highlights: 51.9% of students use iPad for reading less than an hour per day, 81% say that the benefit they got is that it is portable (easy to carry around).

While differences between the two groups of participants were numerous, there were some noteworthy similarities, too. Most notably, both groups enjoyed sharing the content from iPads and collaboration. All interviews with students were semi-structured and one of the questions we asked them all was in what situation have you found the iPad most useful. The answer was invariably related to work in smaller groups, when it was possible to share the information. The second most useful situation was while travelling, both in relation to the longer field work trips and localy. The schoolchildren already had the groups around iPad, the question for them was if there were any problem in sharing. Without exception, they said that sharing was not the problem but fun.

Conclusion

The pilot study conducted at the University level has pointed towards non-acceptance of iPads as a learning platform for Geology students. Various challenges contributed to this situation: from problems with physical space to academic challenge. The two variables that played an important role in the study were perceived ownership and perceived ease of use. A more thorough study is called for in order to better understand the ownership issues, especially in relation to the emerging cloud computing. In contrast, in the elementary school, we observed, but also heard from the families, the children and the teacher, that iPad enhanced teaching, learning and play. The variables that were most prominent for the acceptance of the technology were creativity, attitude toward learning and the emergence of new social patterns. The study with schoolchildren is now continuing, observing if there would be any changes in established patterns with prolonged use, as well as closer observation of how the device actually contributes (or not) to the learning itself now that the novelty of it is in the past.

Acknowledgments

University bookstore and publishing house, for providing the iPads and the University of Oslo Library for project organisation and support.

References

Barki, Henri; Benbasat, Izak. Quo vadis TAM? // *Journal of the Association for Information Systems*, 8 (2007), 4; 244–254

Buckingham David. Beyond Technology: Children's Learning in the Age of Digital Culture. Polity Press, 2007

Chen, Brian. Colleges Dream of Paperless, iPad-centric Education*.* April 5, 2010. http://www.wired.com/gadgetlab/2010/04/iPad-textbooks/ (August 30, 2011)

Culén, Alma Leora; Gasparini Andrea. E-book Reader and the Necessity of Divergence from the Legacy of Paper Book. // *Proceedings of the 4th International Conference on Advances in Computer Human Interaction*. Guadalope (France): (IARIA), 2011, 267 - 273

Culén, Alma Leora; Engen Bård Ketil; Gasparini, Andrea; Herstad, Jo. The Use of iPad in Academic Setting: Ownership Issues in Relation to Technology (Non)Adoption. // *Old meets new – media in education.* Aveiro (Portugal), 2011

Davis, Fred. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. // *MIS Quarterly*, 13 (1989) 3, 319-340

Druin, Alison. Mobile Technology for Children: Designing for Interaction and Learning. Morgan Kaufmann, 2009

Gasparini, Andrea; Culén, Alma Leora. Children´s Journey with iPads in the Classroom. // *Opportunities and Challenges when Designing and Developing with Kids @ School at the Interaction Design for Children Conference* (IDC 2011). Ann Arbor, Michigan, 2011

Green University. Grønt UiO. September 2010.http://www.uio.no/for-ansatte/arbeidsstotte/ prosjekter/gront-uio/prosjektleveranser/rapport-gronn-praksis-final-uten-vedlegg.pdf (July 29, 2011)

Hu, Winnie. More Schools Embrace the iPad as a Learning Tool. January 4, 2011. http://www. nytimes.com/2011/01/05/education/05tablets.html?_r=1 (July 29, 2011)

Hussein, Iltifat; Stanford School of Medicine is giving the iPad to all incoming medical students. July 30, 2010. http://www.imedicalapps.com/2010/07/stanford-school-of-medicine-ipad-incoming-class/ (August 31, 2011)

Sharp, Helene; Rogers, Yvonne; Preece, Jenny. Interaction Design: Beyond Human-Computer Interaction. Wiley, 2007

Staley, David. Adopting Digital Technologies in the Classroom: 10 Assessment Questions, EDUCAUSE 27 (2004), 3

Survey_Geology. iPad Survey Fall 2010. October 2010. http://dl.dropbox.com/u/11591411/ IpadSurvey.pdf (July 29, 2011)

Survey2_Geology. iPad Survey Fall 2010. December, 2010. http://dl.dropbox.com/u/ 11591411/Survey-2-Summary.pdf (August 31, 2011)

Vollen, Dag-Rune. Dansk skole tester Ipad-tvang. 23. August 2011. http:www.idg.no/computerwo rld/article217155.ece?nl=1 (August 31, 2011)

Wilson, Ian. iPads in Education. http://www.iPadineducation. co.uk/iPad_in_Education/Welcome. html (August 31, 2011)

White, Tracie. Will iPad transform Stanford's med school? September 2010. http://med.stanford. edu/ism/2010/september/iPads-0913.html (August 30, 2011)

YouTube. iAnnotate tutorial. Fall 2010. http://www.youtube.com/watch?v=As1Ds 9z8lNI (August 31, 2011)

Cognita's LMS and its Application in the University of Zadar Students' Education

Ljiljana Zekanović-Korona
Department of Tourism and Communication Sciences
University of Zadar
Franje Tuđmana 24i
ljkorona@unizd.hr

Marijana Ražnjević
Department of Tourism and Communication Sciences
University of Zadar
Franje Tuđmana 24i
mraznje@unizd.hr

Božena Krce Miočić
Department of Tourism and Communication Sciences
University of Zadar
Franje Tuđmana 24i
krceb@unizd.hr

Summary

The application of new educational technologies in the education of students and pupils as well as all persons involved in the education system, nowadays becomes an imperative. Cognita's LMS is a programming tool for the creation of electronic educational materials used in distance learning, as well as a supplement to the traditional mode of education.
This paper describes the use of Cognita's LMS as a system for distance learning in students' education. Testing was conducted on a sample of one hundred students of the University of Zadar who have had the opportunity to use the learning system of Cognita.
By means of questionnaire and after using the distance learning system, the students evaluated syllabuses, online tests, the functionality of the system, and communication within the system. This paper examined the quality of educational content created and usefulness of information in this kind of students' education, as well as advantages and disadvantages of this form of education compared to traditional teaching. At the same time, the way students use educational technology was also investigated.

The research data were analyzed by methods of descriptive statistics using the software package of Statistica.
The informatization of education is becoming part of society education and a fundamental prerequisite to better quality education of young people.

Key words: technology, educations, distance learning, Internet, multimedia.

Introduction
In today's post-information age we have been witnessing a widespread use of ICT in all spheres of human activities, including education. Traditional models of education, although still largely present, are replaced or supplemented by new, more modern forms of education supported by the development of computers and computer technology and the Internet. Unlike the traditional model of education, where teachers' knowledge is conveyed in the classroom using chalk, board and language as a medium, more modern forms of education combine information and communication technologies - multimedia and the Internet in order to present knowledge to pupils and students, to interact with them, and to check their knowledge. Users can access online education teaching materials at any time in their normal environment and thus are not forced to attend classes at school.

Many social and economic changes that have marked the time in which we live have also created, besides new forms of education, new forms of learning, so that nowadays the term lifelong learning has become commonly used. According to Jasminka Maravić lifelong learning is defined as "learning activity throughout life, with the aim of improving knowledge, skills and competencies within personal, civic, social and business perspective."[1] A key problem that occurs in lifelong learning is computer literacy, which refers to the ability of individuals being constantly faced with abundance of information to find, extract and use the most needed pieces at a given moment. In addition to information literacy, the user must master new technologies and computer literacy, i.e., to acquire certain knowledge and skills to be able to use efficiently the content offered through the distance learning system.

Distance learning
The historical development of distance learning began more than one hundred years ago, if we include the oldest form of distance education - correspondence education. Distance learning is dependent on the available communication media. The first communication medium for distance education was a letter, and

---

[1] Maravić, J. Cjeloživotno učenje. http://edupoint.carnet.hr/casopis/17/clanci/5.html (accessed on 16 May 2011)

later film, radio and television (public, cable, satellite). With the advent of new communication and computer media, some changes in distance education and the transfer of educational content became possible.

What does the concept of distance learning mean? Distance learning involves a form of teaching and communication between teachers and pupils or students who are separated by a space and sometimes time. By using communication technology, the distance between these two subjects, pupils/students and teachers is overcome, which enables them to communicate directly.

Today, very often the concepts of distance education or online learning are identified with the term of e-learning, which is not the same. E-learning (electronic learning, e-learning) is a form of distance education, but not necessarily be the case. It has been widely used since 1990 and many authors have offered numerous definitions of the term which can be found on the Internet, and in scientific and professional literature. The basic definition of e-learning involves "the use of multimedia and the Internet in order to improve the quality of learning - by providing access to remote information sources and services and facilitating collaboration and communication at a distance." [2]

Depending on the use of ICT in education, there are several forms of e-learning:

- traditional classroom instruction (*face to face*);
- traditional instruction supported by information and communication technology;
- combination of traditional instruction and distance learning, which has wide application in professional seminars and other educational programs;
- online teaching, organized entirely at a distance, using a Learning Management System (LMS) or videoconferencing.

The time and geographic flexibility, financial savings, but also financial benefits are usually mentioned as the main characteristics of e-learning, and thanks to them, e-learning introduces novelties in the education process, opens up new opportunities for exchanging ideas, knowledge and experience to different users. By using e-learning and sitting home at our computer, we meet and connect with experts from all over the world, discussing interesting topics and gaining new knowledge. E-learning is applied in three areas of education: in schools and institutions of higher education, for business education and education in the government sector.

Considering distance learning in higher educational institutions in Croatia, it should be noted that it is often combined with classic forms of education as a mixed model of education.

The introduction of e-learning in educational institutions, particularly higher education ones, according to Dragana Kupres, depends on the commitment and involvement of institutions in the introduction of e-learning at all necessary lev-

---

[2] E-učenje. http://hr.wikipedia.org/wiki/E-učenje (accessed on 16 May 2011)

els.[3] The current problem in Croatia regarding distance education, is still a lack of interest among teachers and pupils / students to use e-learning in the process of teaching in elementary and secondary schools and institutions of higher education. It takes additional training of pupils / students and teachers to apply this form of education through a variety of online courses as those organized and carried out by Croatian Academic and Research Network - CARNet through the E- learning academy and Edupoint courses. It also requires stimulation of teachers for this type of work since making e-learning contents is not part of their jobs. However, pupils or students must, besides the existing skills and habits in a traditional learning, develop new e-learning skills and habits for learning via the Internet.[4]

Distance learning today is often carried out through a Learning Management System (LMS), one of the systems meant for the management of distant learning. LMS has several functions: registration and billing, process management, testing, mentoring and monitoring, user and administrator functions.[5] The first such system in Croatia was developed by Cognita.[6] On the basis of the company's distance learning management system, a survey among students of the University of Zadar was conducted, who, after using Cognita's LMS evaluated courses, online tests, the functionality of the system and communications within the system.

Survey method

The survey was conducted at the beginning of the academic year 2009/10 on a sample of 88 students of the University. The average age of respondents was 20 years.

The survey method comprises completing an anonymous questionnaire with twenty-two questions and circling the multiple choice answers. The results were analyzed using Microsoft Excel and the statistical package of Statistica, version 10.

Results and discussion

The research aimed to determine how respondents use the computer and for what purposes. We were interested in the attitudes of our respondents towards education in the future, and their knowledge of the latest forms of learning. The respondents mainly use computers at home and for more than 5 hours per week.

---

[3] Kupres, D. Obrazovanje za e-learning. http://edupoint.carnet.hr/casopis/24/clanci/2.html (accessed on 16 May 2011)

[4] Watkins. R. Pripremanje studenata za učenje putem Interneta.
http://edupoint.carnet.hr/casopis/42/clanci/2.html (accessed on 19 June 2011)

[5] Löw, Z. Prvi hrvatski LMS. http://edupoint.carnet.hr/casopis/broj-08/clanak-02/index.html (accessed on 18 May 2011)

[6] Cognita company is engaged in the specifics of e-learning and employs specialists in e-learning contents.

Table 1. shows the average marks of the computer use in different software packages. The table shows that the computer is mostly used for writing and editing, and making presentations, followed by more complex programs such as Excel and Access, the programs for image editing and creating web pages and electronic mail.

Table 1. Average marks of the computer use in different software packages

PROGRAMME	AVERAGE MARK
MS Word	4.6
MS PowerPoint	4.6
MS Excel	4
MS Access	3.4
MS FrontPage	4
MS Publisher	2.5
MS Outlook Express	3.3
Adobe Photoshop	3.3
Macromedia Dreamveawer	2.3
Macromedia Flash	2.5
Macromedia Authorware	2.5
Corel Draw	2.7

The analysis of the results showed that all the respondents use only the observed system for e-learning, but have a positive attitude towards lifelong learning and 70% of the respondents plan to expand and update their knowledge in their lifetime. 84% of them think that a good web site is characterized by a good graphic design. 50% of respondents believe that a good web site is characterized by a high quality and clear content, which need not necessarily be current and simple.

Table2. Results of descriptive statistics for evaluation of graphic design, content topicality, content quality, content clarity, content simplicity

VARIABLE	DESCRIPTIVE STATISTICS (SPREADSHEET7)				
	Valid N	Mean	Minimum	Maximum	Std. Dev.
Graphic design	88	3.113636	1.000000	5.000000	0.964089
Content topicality	88	3.386364	1.000000	5.000000	0.836410
Content quality	88	3.068182	1.000000	5.000000	0.968414
Content clarity	88	3.318182	1.000000	5.000000	1.045466
Content simplicity	88	3.136364	1.000000	5.000000	0.846654
Loading speed	88	2.659091	1.000000	5.000000	1.092149

The table 2 gives an overview of the results of descriptive statistics for evaluation of graphic design, content topicality, content quality, content clarity, content simplicity and Cognita's LMS system loading speed. (t-test, df=87)

213

Table 3. Matrix of correlations between the elements of Cognita company's LMS system

VARIABLE	Spearman Rank Order Correlations (Spreadsheet7) MD pairwise deleted Marked correlations are significant at p < 05000					
	Graphic design	Content topicality	Content quality	Content clarity	Content simplicity	Loading speed
Graphic design	1.000000	0.517888	0.518679	0.379489	0.600268	0.660887
Content topicality	0.517888	1.000000	0.554557	0.513669	0.560514	0.517529
Content quality	0.518679	0.554557	1.000000	0.516540	0.525789	0.441803
Content clarity	0.379489	0.513669	0.516540	1.000000	0.418031	0.372836
Content simplicity	0.600268	0.560514	0.525789	0.418031	1.000000	0.642105
Loading speed	0.660887	0.517529	0.441803	0.372836	0.642105	1.000000

It is evident from Table 3 that there is a statistically significant correlation between graphic design, topicality, clarity, simplicity, content quality, and loading speed of Cognita's LMS system. The highest correlation exists between the loading speed of the content and graphic design ($r = 0.660887$, $p < 0.05$).

Table 4. T-test results for the elements of Cognita company's LMS system

VARIABLE	Test of means against reference constant (value) (Spreadsheet 7)							
	Mean	Std. Dev.	N	Std. Err.	Reference constant	t-value	df	p
Graphic design	3.113636	0.964089	88	0.102772	0.00	30.29648	87	0.00
Content topicality	3.386364	0.836410	88	0.089162	0.00	37.98006	87	0.00
Content quality	3.068182	0.968414	88	0.103233	0.00	29.72085	87	0.00
Content clarity	3.318182	1.045466	88	0.111447	0.00	29.77362	87	0.00
Content simplicity	3.136364	0.846654	88	0.090254	0.00	34.75055	87	0.00
Loading speed	2.659091	1.092149	88	0.116424	0.00	22.83981	87	0.00

It is evident from Table 4. that there is a statistically significant correlation between graphic design, topicality, clarity, simplicity, content quality, and loading speed of Cognita's LMS system. The highest correlation exists between the loading speed of the content and graphic design (t-test, df=87).

Table 5. Descriptive statistics results of assessment of navigation and interactivity of Cognita's LMS e-learning system

VARIABLE	DESCRIPTIVE STATISTICS (SPREADSHEET 15)				
	Valid N	Mean	Minimum	Maximum	Std. Dev.
Navigation	88	3.500000	1.000000	5.000000	0.870988
Interactivity	88	3.204545	1.000000	5.000000	0.996074

It is evident from Table 5 that the navigation system for e-learning is assessed at an average grade of 3.5 and the interactivity of the system with 3.2.

Table 6. T-test results for navigation and interactivity of Cognita's LMS system.

VARIABLE	Test of means against reference constant (value) (Spreadsheet 7)							
	Mean	Std. Dev.	N	Std. Err.	Reference constant	t-value	df	p
Interactivity	3.204545	0.996074	88	0.106182	0.00	30.17979	87	0.00
Navigation	3.500000	0.870988	88	0.092848	0.00	37.69615	87	0.00

It is evident from Table 6 that there are significant differences in the assessment of navigation and interactivity of Cognita's LMS system. (t-test, df = 87)

Conclusion

The respondents included in this study used a new way of education i.e. distance learning by using Cognita company's LMS system. We have found that they have not used such methods of education, but have a positive attitude about future education and in their future lifelong learning they are willing to use some parts of the system for e-learning. Graphic design of the e-learning system and the design of teaching contents have an important role in the perception of the system for distance learning. A good system for e-learning, beside a good graphic design, is characterized by good quality, clear, timely and simple contents and loading speed. Graphic design, loading speed and quality, clarity, topicality and simplicity of contents are interrelated and affect the quality of the system for e-learning. The research results suggest the need for greater interactivity of teaching contents and the improvement of the navigation system within the e-learning system. These research results can be useful to all the designers of teaching contents for distance learning. It is a guideline on what needs to be paid attention to when creating high-quality instructional contents within the system for e-learning.

References

E-učenje. http://hr.wikipedia.org/wiki/E-učenje (accessed on 16 May 2011)

Kupres, D.: Obrazovanje za e-learning. http://edupoint.carnet.hr/casopis/24/clanci/2.html (accessed on 16 May 2011)

Löw, Z. Prvi hrvatski LMS. http://edupoint.carnet.hr/casopis/broj-08/clanak-02/index.html (accessed on 18 May 2011)

Maravić, J. :Cjeloživotno učenje. http://edupoint.carnet.hr/casopis/17/clanci/5.html (accessed in May 2011)

Watkins. R.: Pripremanje studenata za učenje putem Interneta. http://edupoint.carnet.hr/casopis/42/clanci/2.html (accessed on 19 June 2011)

E-Learning in LIS Education: Satisfaction of Part-Time LIS Students with Omega

Marko Tot
Faculty of Humanities and Social Sciences
Ivana Lučića 3, Zagreb, Croatia
mtot@ffzg.hr

Summary

The application of e-learning in Croatian universities has increased rapidly, with the introduction of the Bologna process to create the European Higher Education Area. The application of digital media for teaching and learning makes distance education for LIS professionals at the Faculty of Humanities and Social Sciences, University of Zagreb, possible. E-learning and traditional classroom learning have been combined to deliver library and information science (LIS) education. The aim of our research was to obtain a general overview of the part–time LIS students' expectations and experiences using Omega, a specific learning management system.

Key words: e-learning, LIS education

Introduction

Implementation of ICT into education and training was encouraged as early as 2001 by the European Council Resolution on e-Learning. It invites the Commission to pay particular attention to the implementation of the e-Learning action plan and to the concrete future objectives of education and training systems, in line with the objective set by the Lisbon European Council to make the EU the world's most competitive and dynamic knowledge-based economy by 2010.[1] According to a UNESCO report "Open and distance learning is one of the most rapidly growing fields of education, and its potential impact on all education delivery systems has been greatly accentuated through the development of Internet – based information technologies, and in particular the World Wide Web."[2] E-learning definition states that it is learning that is delivered, enabled or mediated by electronic technology, for the explicit purposes of training

---

[1] Council Resolution on e-Learning. Europa: summaries of EU legislation. http://europa.eu/legislation_summaries/education_training_youth/lifelong_learning/c11052_en.htm (28.6.2011.)

[2] Moore, M. M.; Tait, A. Open and Distance Learning – Trends, Policy and Strategy Considerations, UNESCO Division of Higher Education, Paris 2002. p. 7 http://unesdoc.unesco.org/iimages/0012/001284/128463e.pdf (28.6.2011.)

and/or education. It does not include stand-alone technology-based training such as the use of CD-ROMs in isolation.[3] E-learning has, therefore, become an important component of both formal and non-formal education. The advancement of e-learning in universities is also influenced by the introduction of learning management systems. In addition to ICT initiatives, terms like learner-oriented teaching, learning outcomes and satisfaction are being mentioned in the academic life both by students and teachers more and more frequently.

In 2007, the University of Zagreb drafted the E-learning Strategy 2007 – 2010, which states that "the University of Zagreb supports and actively promotes e-learning and the application of information and communication technology in teaching and learning at all levels of university education. E-learning is a legitimate and desirable way of learning and teaching at the University of Zagreb."[4] The Strategy points out that the four goals the University of Zagreb aims at achieving, through e-learning, are the following:

- to improve the quality of university education
- to enable teachers and students to play new roles in the educational process
- to increase the competitiveness of the University and university curricula
- to train students to use technology for lifelong learning.

The open-source course management system Moodle has been used in the Faculty of Humanities and Social Sciences, of the University of Zagreb, since the academic year 2000/2001. A survey conducted in 2007 examined students' attitudes and satisfaction with the use of Omega, a customized version of Moodle in use since the year 2004/2005. The majority of the students were from the Departments of English (34%), of Sociology (29%), and of Information Sciences (20%). The results indicated positive attitudes towards Omega.[5] The first research among LIS students' carried out in the Department of Information Sciences showed also that students have a positive attitude towards Omega. This research also showed that Omega makes studying easier.[6]

---

[3] Allan, B. E-learning and teaching in library and information services. London: Facet publishing, 2002.

[4] University of Zagreb E-learning Strategy 2007. – 2010. http://www.unizg.hr/fileadmin/rektorat/dokumenti/eucenje_strategija/University_of_Zagreb-E-learning_strategy.pdf (28.6.2011.)

[5] Klasnić, K; Seljan, S; Stančić, H. Quality parameters for the e-learning Omega system. http://bib.irb.hr/datoteka/361763.Klasnic_Seljan_Stancic.pdf (28.6.2011.)

[6] Tot, M. ; Živković, D. The role of e-learning in LIS education: students' evaluations. // MIPRO 2011 / 34th International Convention Computers in Education, May 23-27, 2011 / Čičin-Šain, Marina (ed.). Rijeka : MIPRO, 2011. p. 263-266.

The aim of this paper is to see what the attitude of the part-time students towards Omega is. The reason why the author decided to conduct a survey is also to make comparison between the full-time and the part-time LIS students.

Research

A research was carried out among the part-time students of library and information science of the Faculty of Humanities and Social Sciences in Zagreb. The research included part-time students who attended the Reference Work in Libraries course. The survey questionnaire was not completed by all the attendants of the students, since they were not present at the lecture, for valid reasons. Still, the research sample was representative because it was carried out on a majority of the students. Out of the 90 students enrolled in the course 67 students were present in class and they all completed the survey. These are the students who have already completed another course of study, and are currently employed in a library. Since they have no formal education in the field of library and information science, in line with the Croatian Library Act they are required to complete a two-year study of library and information science while working full time. Part-time students didn't have any official training on how to use Omega, but every professor introduces students with his course on Omega. Also every professor decides if usage of Omega is requirement for passing the course. At Reference Work in Libraries course usage of Omega wasn't obligatory to pass the course. The aim of the research was to obtain a general overview of the part-time students' satisfaction with Omega, a distance learning tool, and to establish to what degree this tool makes studying easier for students; and also to compare the results obtained in the present research with the results of the previous research conducted among full-time students of library and information science. The authors assumed that Omega met the students' needs and helped making their study much easier.

Methodology

The part-time students at the Reference Work in Libraries course were asked to complete an anonymous survey questionnaire, consisting of 13 questions. There were only two questions, out of the 13, that required from students to write their own answers, other questions provided multiple choice answers for students to choose from. The first three questions were general questions aimed at getting data about students' gender, age and their so-far achieved education. Questions 4 to 9 were essential ones related to Omega: is it used, how often, how many courses the student has on it, is this an e-learning tool, can Omega replace lectures. Questions 10 to 13 evaluated students' satisfaction with Omega, providing also answers about how much it makes studying easier, how satisfied they are with the content offered, and what is on Omega used by the students most, using grading 1 to 5, where 1 indicated the lowest grade of satisfaction, and 5 the highest.

Results

The total of 67 students completed the survey, out of which 13% male and 87% female students; 24% respondents were between 24 and 27 years of age, 18% between 28 and 30 years of age, 42% between 31 and 41 years of age, and 16% between 41 and 50 years of age. All the respondents replied that they used Omega: to the question about how often it was used, 14% replied every day, 46% replied 2-3 times a week, 37% replied 2-3 times per month, and 3% replied that they used Omega very rarely. Asked how many courses the respondents had included in Omega, 4% of the students replied they had 1 – 3 courses included, 30% had 3 – 5 courses included, 45 respondents had most of the courses included in Omega, 19% had all the courses included in Omega, and 2% did not provide an answer to this question. Asked whether the respondents were familiar with the e-learning concept, 93% of them replied YES, and 7% replied NO. Omega is considered an e-learning tool by 97% of the respondents, and 3% of them do not consider it an e-learning tool. It is considered that Omega may replace lectures by 34% of the students, 61% consider that Omega cannot replace lectures, and 5% did not provide an answer to this question. The respondents' satisfaction with Omega was distributed as follows: 24% of the respondents evaluated Omega as excellent, 48% as very good, 24% as good, 3% as satisfactory, and 1% of the respondents evaluated their satisfaction with Omega as negative. Not one of them evaluated their satisfaction with Omega as negative. The satisfaction with Omega's contents was evaluated as follows: 27% of the respondents evaluated the contents as excellent, 45% as very good, 22% as good, and 6% as satisfactory. The respondents evaluated how much Omega makes their study easier, as follows: 43% of the respondents evaluated Omega in this respect as excellent, 30% as very good, 24% as good, and 2% as satisfactory, and 1% as unsatisfactory.

Table 1: Satisfaction with Omega.

	5	**4**	**3**	**2**	**1**
satisfaction with Omega	24%	48%	24%	3%	1%
Omega is making studying easier	43%	30%	24%	2%	1%
satisfaction with content on Omega	27%	45%	22%	6%	0%

In the last question the respondents evaluated what contents they use the most on Omega with grades from 1 to 5. The contents offered were the following: news forum, course description and reading list, presentations used in lectures, additional material related to the course, articles from the Faculty repository, submission of home-works and seminar papers, communication with the teaching staff, communication with students.

220

Table 2: Usage of content in Omega.

	5	**4**	**3**	**2**	**1**	**no answer**
news forum	34%	21%	9%	14%	19%	3%
course description and reading list	64%	27%	4%	3%	0%	2%
presentations used in lectures	82%	11%	4%	1%	0%	2%
additional material related to the course	27%	28%	27%	9%	7%	2%
articles from the Faculty repository	24%	24%	25%	11%	13%	3%
submission of home-works and seminar papers	19%	14%	22%	12%	28%	5%
communication with the teaching staff	17%	22%	15%	18%	22%	6%
communication with students	7%	15%	9%	9%	54%	6%

Their use of news forum was graded as excellent by 34% of the respondents, by 21% as very good, by 9% as good, by 14% as satisfactory, by 19% as unsatisfactory, and 3% did not evaluate how much they used the news forum. Their use of course descriptions and the reading lists was graded as excellent by 64% of the respondents, by 27% as very good, by 4% as good, by 3% as satisfactory, none of the respondents graded their use of course descriptions and the reading lists as unsatisfactory, and 2% of the respondents provided no answer to this question. The use of presentations from the lectures was graded as excellent by 82% of the respondents, by 11% as very good, by 4% as good, by 1% as satisfactory, and 2% of the respondents provided no evaluation. The use of additional material was graded as excellent by 27% of the students, by 28% as very good, by 27% as good, by 9% as satisfactory, by 7% as unsatisfactory, and 2% of the respondents provided no evaluation. The use of articles from the Faculty repository was graded as excellent by 24% of the students, by 24% as very good, by 25% as good, by 11% as satisfactory, by 13% as unsatisfactory, and 3% of the respondents provided no evaluation. Using the possibility to submit home-works and seminar papers via Omega was graded as excellent by 19% of the respondents, by 14% as very good, by 22% as good, by 12% as satisfactory, by 28% as unsatisfactory, and 5% of the respondents provided no evaluation. Using the communication with the teaching staff was graded as excellent by 17% of the respondents, by 22% as very good, by 15% as good, by 18% as satisfactory; 22% of the respondents graded their using the communication with the teaching staff as unsatisfactory, and 6% of the respondents provided no evaluation of their using the communication with the teaching staff. Using the communication with students was graded as excellent by 7% of the respondents, by 15% as very good, by 9% as good, also by 9% as satisfactory; 54% of the respondents graded their using the communication with the students as unsatisfactory, and 6% of the respondents provided no answer to this question.

Discussion

The results obtained show that the part-time students use the distance learning tool Omega, and that high percentage of them uses such tool relatively often (14% use Omega every day, and 46% use Omega 2-3 times a week). Even so, the number of 37% of the students using Omega only 2 – 3 times a month causes concern. One of the reasons for such use may be also the fact that 30% of the respondents have 3 – 5 courses included in Omega. However, 45 % of the students have most of the courses included in Omega, 19% of them have all the courses in Omega, so that these results show that e-learning has been accepted as a standard part of tuition. The survey shows that the students are familiar with the concept of e-learning, and they consider it supplementary to traditional learning (61% of the students believe that Omega cannot replace lecturing). Based on the research results, we can determine that the students are satisfied with Omega (24% evaluated Omega as excellent, and 48% as very good), and that Omega makes studying easier (43% evaluated Omega in this respect as excellent, and 30% as very good). Yet, one should not disregard that 1% of the students evaluated their satisfaction with Omega negatively, and that the use of Omega to make studying easier was evaluated negatively by 1% of the students, which may be a consequence of the fact that a part of the students do not use Omega often. The satisfaction with Omega's contents was not evaluated negatively by any of the respondents. This shows that the students should be encouraged to use Omega more, so that Omega would make their studying easier. The contents found in Omega and used most often by the part-time students are the following: presentations used in lectures, course descriptions and reading lists, additional materials related to the course and news forum. The contents in Omega used least often by the students are communication with their fellow students and communication with the teaching staff. This may be caused by the fact that they still use other media for communicating with the teaching staff and their fellow students, such as telephone and e-mail, so that the communication via Omega is not needed.

The author carried out the same research also among full-time library and information science students, and those results were presented at MIPRO 2011 conference.[7] The sample used for the research on regular students was 68 students, thus making these two samples very similar so that they can be compared. However, it is necessary to point out that the part-time students completed their first course of study before Omega existed, and a considerable

---

[7] Tot, M; Živković, D. The role of e-learning in LIS education: students' evaluations // MIPRO 2011 / 34th International Convention Computers in Education, May 23-27, 2011 / Čičin-Šain, Marina (ed.). Rijeka : MIPRO, 2011. p. 263-266.

number of them completed their first course of study even before internet existed. As the results of the survey show, most of the respondents (42%) are between 31 and 40 years of age, and there were also older respondents (16% between 41 and 50 years of age), and they are only now coming to grips with computer literacy. Having compared the results of the full-time students and the part-time students of library and information science, it is evident that the full-time students use Omega more. The reason for this may be that the part-time students are employed, and they have less time that the full-time students, while the full-time students belong to the Google Generation, and spend most of their time on internet. Both groups of students are satisfied with Omega, but higher percentage of part-time students evaluated their satisfaction with Omega as being excellent (24% of the part-time students in contrast to 13% of the full-time students). Likewise, being asked how much Omega makes their studying easier, higher percentage of the part-time students evaluated this tool as excellent (43% of the part-time students in contrast to 28% of the full-time students), but the total number of students, who evaluated Omega as excellent and very good in making their studying easier, is higher among full-time students (78% of the full-time students evaluated it as excellent and very good, and 73% of the part-time students evaluated it as excellent and very good). Higher percentage of the full-time students are familiar with the e-learning concept (94% of the full-time students, and 93% of the part-time students), and higher percentage of the part-time students believe that Omega is an e-learning tool than the full-time students (97% of the part-time students, 91% of the full-time students).

Table 3: Part-time and full-time students satisfaction with Omega.

		5	4	3	2	1
FULL-TIME STUDENTS	satisfaction with Omega	13%	58%	28%	1%	0%
	Omega is making studying easier	28%	50%	22%	0%	0%
	satisfaction with content on Omega	9%	57%	30%	4%	0%
PART-TIME STUDENTS	satisfaction with Omega	24%	48%	24%	3%	1%
	Omega is making studying easier	43%	30%	24%	2%	1%
	satisfaction with content on Omega	27%	45%	22%	6%	0%

Conclusion

This research shows that e-learning has become integrated in LIS curriculum in the Faculty of Humanities and Social Sciences in Zagreb. A tool for e-learning in the Faculty is called Omega. A survey was conducted among part-time students to see what their opinion on Omega was and whether Omega was making their studying easier. The results show that students often use Omega, that they have the majority of the courses on Omega. The survey also shows that the

majority of students are familiar with the concept of e-learning, considering it supplementary to traditional learning. The students are satisfied with Omega, and Omega makes their studying easier. The contents used most often by the students are the following: presentations used in lectures, course descriptions and reading lists, additional materials related to the course, news forum. Comparing regular and part-time students, the survey has shown that full-time students use Omega more often. Both groups of students are satisfied with Omega and Omega is making their studying easier. However a higher percentage of part-time students assessed Omega excellent. But total number of full-time students who evaluated that Omega makes their studying easier with excellent and very good is higher than the number of part-time students. The author believes that this type of research should be conducted regularly on annual basis, thus enabling e-learning to be improved.

References

Allan, B. E-learning and teaching in library and information services. London: Facet publishing, 2002.

Council Resolution on e-Learning. Europa: summaries of EU legislation. http://europa.eu/legislation_summaries/education_training_youth/lifelong_learning/c11052_en.htm (28.6.2011.)

Klasnić, K; Seljan, S; Stančić, H. Quality parameters for the e-learning Omega system. http://bib.irb.hr/datoteka/361763.Klasnic_Seljan_Stancic.pdf (28.6.2011.)

Moore, M. M.; Tait, A. Open and Distance Learning – Trends, Policy and Strategy Considerations, UNESCO Division of Higher Education, Paris 2002. p. 7 http://unesdoc.unesco.org/images/0012/001284/128463e.pdf (28.6.2011.)

Tot, M.; Živković, D. The role of e-learning in LIS education: students' evaluations. // MIPRO 2011 / 34th International Convention Computers in Education, May 23-27, 2011 / Čičin-Šain, Marina (ed.). Rijeka : MIPRO, 2011. p. 263-266.

University of Zagreb E-learning Strategy 2007. – 2010. http://www.unizg.hr/fileadmin/rektorat/dokumenti/eucenje_strategija/University_of_Zagreb-E-learning_strategy.pdf (28.6.2011.)

Framework of the Language Learning Environment for Assisting Foreigners in Learning Croatian (AFILC)

Kristina Posavec
Faculty of Humanities and Social Sciences
Ivana Lučića 3, Zagreb, Croatia
kristina.posavec@gmail.com

Nives Mikelić Preradović
Faculty of Humanities and Social Sciences
Ivana Lučića 3, Zagreb, Croatia
nmikelic@ffzg.hr

Sanja Kišiček
Faculty of Humanities and Social Sciences
Ivana Lučića 3, Zagreb, Croatia
smatic@ffzg.hr

Summary

The goal of this paper is to present the development framework of the interactive multimedia project AFILC (Assisting Foreigners in Learning Croatian). The purpose of the AFILC is to develop language learning materials that will motivate foreigners to learn the Croatian language and help them cope with its grammatical richness. Interesting graphics, motivating user-friendly interface, educational character and interactivity are the characteristics that will enhance the process of mastering the Croatian language for foreigners. In the AFILC the users will be able to almost completely control and manage the process of learning at their own pace.

Key words: e-learning, computer assisted language learning (CALL), the Croatian language, multimedia, foreigners, education

Introduction
In this paper we will present the concept of the AFILC - the interactive multimedia learning system that will make the process of learning Croatian as a foreign language easier and more appealing to foreigners. The main idea is to make learning more dynamic and appealing to the specific group of people (adult learners between ages 25 and 40) who want to learn the Croatian language. The AFILC system will be available online and free of charge, so that

each user has the opportunity to access and manage all the learning materials. The application interface will be interactive and all the users will be able to follow their own learning path, hopefully improving their language skills. Also, one of the main characteristics of the AFILC system will be its communicative approach, focusing on the language use rather than the language analysis.

However, foreign language learning is not the only aim of this project. Our secondary goal is to link the process of language learning with a wider knowledge of Croatian culture, heritage and history of the Croatian people. All educational materials will consist of carefully selected texts about Croatian history, literature, poetry, art and tradition as well as contemporary issues. Also, we will include facts about some of the most famous Croats that are recognized all over the globe, for example Eduard Slavoljub Penkala, Janica Kostelić, etc.

Description of the AFILC system

The software for learning Croatian as a foreign language will be designed for beginning level of learning. The target group refers to people between ages 25 and 40 who are information technology (IT) literate. The requirement of IT literacy is important because the software will be interactive and it will require the user to manage all the included material. The content will be based on everyday communication in Croatian. Communicative language learning is more appropriate for performing individual beginning level lessons performed by the help of a computer application than analytical learning [13], which is based on studying the morphological structure of the language. The analytical method is appropriate for higher levels of language learning, such as intermediate, advanced and professional. We believe that beginner users should learn basic concepts in a foreign language, and build their own dictionary, which will enable them to communicate in Croatian and cope with everyday situations. The basic language that will guide the user through the content in the software will be English. Application is intended for foreigners who do not belong to Slavic nations, because the Slavs are not absolute beginners in the Croatian language due to similar language structures. Initial version of the AFILC software will offer 20-30 words in a dictionary, and between 10 and 15 most common communication phrases. Based on presented words and phrases, we will provide an explanation on discrepancy between nouns, verbs, adjectives, numbers (1-10) and pronouns (personal and possessive). The last two sections will apply on learning the three most common verb forms (simple present tense, simple past tense and simple future tense), and two cases with their corresponding declensions, i.e. noun inflections (nominative and accusative case). We will put emphasis on practical work and exercises, because we believe this kind of approach is most appropriate for computer assisted language learning (CALL). We also believe that interactivity has a greater impact on users' perception and adoption of language than classical education based on textbooks.

Learning modalities and cognitive styles in the AFILC system

Learning modalities are sensory based and refer to the primary way our bodies take in information though our senses: visual, auditory, kinesthetic and tactile [8].

On the other hand, cognitive styles represent a consistent approach to organizing and processing information, typical modes of problem solving, thinking, perceiving and remembering. Global style learner is one who processes information and sees the perceptual field as a whole, more socially oriented person, while the analytic style learner can easily break the field down into its component parts and is not influenced by the existing structure.

The people from our target group need a constant link between the learning content and real life situations because it is very important for them to quickly adapt in everyday life. Therefore, they are global style learners [7]. The main feature of the global cognitive style is the establishment of interaction between the user and the language used in everyday communication. Instead of classical learning, where a user gets a textual hint in the parenthesis, we will use graphical material as we can see in example 1. Combining global cognitive style with visual and auditory learning modalities our users will utilize the AFILC as interactive multimedia software for learning. The visual modality is characterized by use of words and phrases that evoke visual images, while the auditory modality pertains to thinking in a linear manner, listening and verbalizing words [7]. While using the AFILC, users will be able to hear the pronunciation of certain words, phrases or even complete sentences in the Croatian language and that will make learning easier. We will give examples of useful phrases for getting around in space and time, for example "Where is the Botanical Garden?", "How to get to the Museum of Zagreb?" or "Do you know what the time is?" We believe this way the users will master some basics of everyday communication. The application will include a dictionary in which every word will contain a textual explanation both in English and in Croatian. The pronunciation will be only in Croatian. Auditory based exercises will be designed in the form of dictates. User will have to enter the text that is dictated to him into a certain field. At the end, a user will get the correct answer, i.e. properly written text, and an insight into mistakes. The visual modality will be encouraged by a combination of text and pictures, as presented in example 1. Vocabulary exercises will include graphical material where the user will have to recognize what the image shows, for example, exercise of identifying and writing words, as it is shown in example 2. The AFILC will cover orthographic learning, so that users learn how to write words and orthoepic learning in order to hear the correct pronunciation of a given word. Each example will be followed by a pronunciation of the word in the whole sentence. When learning vocabulary, the semantic level is very important, since it explains the word that the user is learning. This explanation will be available in two languages, both English and Croatian. The pragmatic aspect of learning is also important, since the AFILC will allow the user to learn

the basic phrases that are used in everyday communication, which will enable the user to learn about the general culture of Croatia (example 3).

Example 1
Fill in the blanks. The picture might help you.

Koji _____ vozi do Trga bana Josipa Jelačića?

Trenutno je _____ sati.

Example 2
Drag the image of the peach in the circle and then write the correct word in the text field and click on the check-box button to get the correct answer.

Multimedia and interactivity in the AFILC project

The AFILC project is based on several multimedia principles that will be explained within this chapter. The learning content will be presented to users in a dynamic way that includes picture, sound, video and animation in order to make the language learning process more efficient and easier than it is in the traditional classroom environment [1, 11].

We will use animation moderately, since the excessive use of animation can be a distraction for users and change their focus [3, 10]. Usage of colors will be balanced and not too intensive. The audio materials will be clear and comprehensible, without any background noise, while the narrator voice will be pleasant and calming with an emphasis on the parts that are important. Finally, our aim is to create a balanced relationship between image, sound and text in the application [5].

The AFILC system will use controlled interactivity that enables user to individually navigate through the content. The navigation bar will allow user to control the displayed content and navigate from one unit to another in a simple and easy way.

228

The idea of AFILC isn't to replace language teachers; the idea is rather to improve the users' learning skills enabling them to master the basics of the Croatian language quickly and more efficiently using well-designed multimedia activities that want to imply multimedia principles and interactivity in language learning situations.

Quizzing in the AFILC system

As it was stated before, adult foreigners that are beginners in the Croatian language will make the target group of users using the AFILC. Therefore, the test will be created for beginning level of knowledge. The tests will not be time limited and users will be able to work with the exercises at their own pace. There will also be the possibility of self-evaluation after solving the exercises. Upon completing a given task, users will see the correct answers immediately and get an insight into their mistakes. Knowledge testing will be conducted in a form of a quiz. The following question types will be used: choice, matching, filling in the blanks and short answers. Essay questions will not be used since they are "reserved for spoken and written production"[6] and since an individual learning a foreign language for the first time cannot perform self-evaluation on this type of question. The primary purpose of the AFILC software is acquiring basic knowledge of the Croatian language. An example of an educational activity for knowledge assessment that will be used in the software is given below. Explained example can be graphically seen in Example 3:

Description of activities
Unit: general knowledge related to Croatian culture.
Communication themes: vocabulary knowledge check, understanding the questions, mastering general culture.
Learning content: nouns, verbs, pronouns, numbers, colors
Mode: individual
Degree: beginner level

Purpose of the educational activity: acquiring Croatian vocabulary and proper usage of the terms in a given context. The aim is that users master correct writing and adopt general Croatian culture.

Exercise description: the interactive crossword appears on the computer screen with fields that users have to fill in. Below the crossword, there will be questions in English. Questions and fields in the crossword will be matched by the same number. Answers to the questions will correspond to fields in the crossword. At the end of each line in the crossword puzzle there will be an interactive button. After entering a term, users can click on the button to get the feedback on the correctness of their answer. Exercises can be repeated as many times as necessary.

Example 3:
Read the questions and fill in the crossword puzzle.

1	z	a	g	r	e	b

Crossword puzzle grid:

1	z	a	g	r	e	b					
2	o	s	a	M							
3	o	l	o	v	k	a					
4	s	k	i	j	a	n	j	e			
5	p	l	a	v	A						
6	j	a	d	r	a	n	s	k	o	-	m	o	r	e
7	l	i	j	e	p	a	-	n	a	š	a
8	k	r	a	v	a	t	a

1. Which is the capital of Croatia?
2. How many national parks are there in Croatia?
3. What was invented by Slavoljub Eduard Penkala?
4. In which sport did Janica Kostelić win four golden Olympic medals?
5. Which color is the trademark of the famous Zagreb football club Dinamo?
6. What is the name of the Croatian sea?
7. What is the name of the Croatian anthem?
8. What is the name of a decorative piece of clothing in the form of ribbon that is tied around the neck?

Issues and possible problems

Each project meets uncertainties in design and implementation, as well as in methods that will be presented for the purpose of acquiring knowledge. What we want to avoid in the AFILC system is uniformity of the content. We will try to present the content in a dynamic way by implementing interactive activities such as crossword puzzles (example 3). Furthermore, we want to have contextually dependent materials and articles. We will use authentic text documents in order to reach two basic goals: firstly, we want to give examples of a basic everyday communication, and secondly, we want to introduce contemporary themes concerning Croatia to strangers in order for them to get better acquainted with Croatian culture because "if you study the language, while not studying the culture within which the language works, it carries only meaningless symbols or symbols that students attribute to the wrong interpretation" [4]. A very important issue of computer-assisted (foreign) language learning is its accessibility which makes it different from learning in a traditional way, by using books within a student-teacher relationship [9]. Textbooks and classroom teaching are not available to all users, while a software application published on the Internet can be available to larger masses at any place any time. Another

important issue is the motivation of users, which we tend to increase and sustain by the inclusion of dynamic games and exercises (examples given). In developing the AFILC, we will use Adobe Flash CS5, a comprehensive tool which enables designing software with a considerable amount of interactivity for dynamic learning. We will not use the software that does not support the diacritical characters, which are an integral part of the Croatian language. We will not use software such as Easy generator, Dreamweaver or Microsoft Office Power-Point, which is commonly used in drill and practice exercises. The reason for not using PowerPoint is that its' primary purpose is presentation, rather than interactivity. We also have to consider the size of the application for two reasons: firstly, for storing the application on a portable media, such as a DVD or a CD ROM, and secondly, so it can be run on older and/or slower computers. Therefore, the graphics in the software should not be too demanding. We wish to create the AFILC software that will be available online so that users can access it from all over the world. Since there are plenty of computer training programs for preschool and school age (e.g. *Zekina košarica, Sunčica među slovima, među brojka, u prirodi i u prometu*), we plan for the AFILC to be a computer education program for users between ages 25 and 40 that have basic ICT literacy. Therefore, the design and usage has to be adapted to those structures. We will avoid too colorful content, as well as too cheerful music, interface congestion with unnecessary animations and graphics that distract from the main purpose, i.e. learning. Instead, we will use video clips, photographs, illustrations, moderate animations and pronunciation. The content will include articles intended for the target group (music genre, trends, fashion, etc.), phrases for getting around in the city and concepts from general Croatian culture (sports, film, art, nature and cultural heritage). We want to follow a constructivist approach of learning with a learner at the centre of the educational process [2]. The AFILC software will be useful, cheaper and more attractive compared to learning from textbooks. Although the main goal of this project is learning a foreign language, we must not neglect the second most important goal, which is the integration of learners in the Croatian society through communication and contemporary themes.

Lexical issues of cases and verb forms
The most common problem in mastering the Croatian language is production of meaningful sentences using appropriate verb forms and cases. Since this project is for beginning level of learning, our wish is to create a good foundation for later learning process so that the user could better master the higher levels of the Croatian language. We will concentrate on mastering the basic verb forms only; simple present, simple past and simple future tense because we believe those are necessary for basic communication. We will present only two cases, nominative and accusative, out of seven altogether. We will give examples of noun inflection for the two cases so the users gain basic knowledge for further understand-

ing of the rest of the cases and corresponding inflections. The reason why we chose the nominative is because it is a "neutral case, used for the nomination and accusative is taught second since it is the most common case after the nominative case and has a simple meaning that can be a good learning material in beginner level communication" [12]. We will try to resolve the problem of acquiring different verb forms and cases by forming thematic units that will be used for basic communication phrases for getting around in space and time. In the unit with verb forms, we will initially include adverbs of time at the beginning of sentences in order to help users to master simple present, simple past and simple future. Afterwards we will form sentences without adverbs of time. Together with verb forms we will introduce accusative case. For beginning level, it is important that users know to which questions the cases respond to. This would ease the proper use of cases. Nominative case will be demonstrated on the example of people introducing each other (examples 1, 2) or on explanation of words and concepts (3, 4). Learning the accusative case by the help of a verb that can be followed by a direct object, such as eat, drink or look will clarify the difference between nominative and accusative case (5, 6, 7). In addition to the textual content, pronunciation of words or whole sentences is essential in order for the user to hear the difference between the nominative and accusative case, which will be emphasized all the time.

Example 4:
    (1) Moje ime je Marko. → tko? Marko
    (2) Ja sam Marko. → tko? Marko

    (3) Ovo je restoran. →što? restoran
    (4) Ovo je ptica. →što? ptica

    (5) Jedem ribu.
    (6) Jest ću ribu.
    (7) Jeo sam ribu.

The software will support pronunciation of both incorrect and correct answers. This option will enable the user to hear the difference between correct and incorrect answers. Verb pronunciation will be accompanied by a video clip presenting the action of the verb. A user will have to recognize what action is presented after which he will have to provide the correct verb form in simple present, simple past and simple future tense.

In example 4 there are exercises for noun inflections in the nominative and accusative case. By the help of photos, a user will have to fill in the blanks in the sentences. After filling in the blank, a user clicks the check-box button on the right side and the software provides feedback on the correctness of the answer. In either case, a user will get an explanation why the answer is correct or why it

is incorrect. After that the user will be able to click on the button for pronunciation of the sentence, which will enable hearing the difference between the nominative and accusative case. The goal of this type of exercise is that the user understands the differences between the nominative and accusative case both in terms of grammar and pronunciation. Grammar definitions and inflection rules are not sufficient for successful language learning. Through interactive exercises, video clips, animation, and pronunciation, users will master Croatian grammar and spelling much quicker and easier than by using textbooks. The AFILC software will provide exercises and explanations, and users will be provided by feedback so they learn on their mistakes. The educational content will not be comprehensive, since the program is designed for beginning level of language learning.

Example 5

Jedem (što?/what?) [_____]   ⬅🐟   ☑   🔊

Ovo je (što?/what?) [_____]   ⬅🐟   ☑   🔊
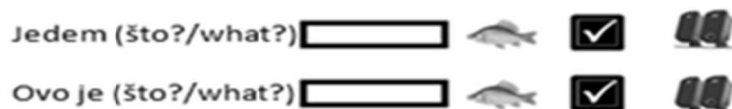
Evaluation and conclusion

The evaluation of the AFILC system will involve the testing of the effectiveness and usability. In order to measure the effectiveness of the software, we will conduct a research involving two groups of learners, the control group and the experimental group of learners. Users in the experimental group will use the AFILC as a learning instrument, while the users in the control group will use the conventional learning method. A pre-test and a post test will be conducted with both groups. By means of a pre-test we will see if all the users are on the same level of knowledge in Croatian. We will also test the users' information literacy. The questions will be designed to assess the users' understanding of the specific language lesson. The results of these tests will be compared to measure the users' performance.

The usability evaluation will be performed by distributing a set of questionnaires to the experimental group after they have learned a specific lesson using the AFILC. Four usability factors will be used to evaluate the system: learnability, efficiency, screen design and satisfaction. Those factors will also be used to evaluate the usability of the specific multimedia elements integrated into the system.

References

Abhaya Asthana. Multimedia in Education - Introduction, The Elements of, Educational Requirements, Classroom Architecture and Resources. http://encyclopedia.jrank.org/articles/pages/6821/Multimedia-in-Education.html (19th May 2011)

Alessi, Stephen M, Trollip, Stanley R. Multimedia for learning. Needham Heights: Alyn and Bacon; 2001.

Betrancourt, Mireille. The Animation and Interactivity Principles in Multimedia Learning.//The Cambridge Handbook of Multimedia Learning. Cambridge University Press. University of California. Santa Barbara. 2005.

Bratanić, Maja. (1993) Nastavnik stranog jezika kao tumač strane kulture.// Trenutak sadašnjosti u učenju jezika. Zagreb: Hrvatsko društvo za primijenjenu lingvistiku.

C.W. Debuse, Justin; Hede, Andrew; Lawley, Meredith. Learning efficacy of simulataneous audio and on-screen text in online lectures. // Australasian Journal of Educational Technology. Vol. 25, No. 5 (2009), 748-762

Ferbežar, Ina; Požgaj-Hadži, Vesna. Kako izgraditi jezični test.//Lahor. Vol 2(2008), No.6; 165-182

Fowler, B.T. The effectiveness of computer-controlled videodisc based training. Unpublished doctoral dissertation, University of Iowa, Iowa City, 1980

Grgić, Ana; Kolaković, Zrinka. Primjena stilova i nastavnih strategija u nastavi hrvatskog kao inog jezika.//Lahor. Vol.1(2010), No.9;78-96

Levy, Michael. Computer-assisted language learning: context and conceptualization. Oxford: Oxford University Press, 1997

Mayer, Richard E.; Moreno, R. (1998). "A Cognitive Theory of Multimedia Learning: Implications for Design Principles". http://www.unm.edu/~moreno/PDFS/chi.pdf. (17th May 2011)

Najjar, Lawrence J. Multimedia Information and Learning.// JI. of Educational Multimedia and Hypermedia Vol. 5 (1996), No. 2, 129-150

Udier, Sandra Lucija; Gulešić-Machata, Milvia; Čilaš-Mikulić, Marica. Gramatičko-semantički pristup obradi padeža.//Lahor. Vol.1(2006), No.1; 36-48

Underwood J. Linguistics, computers and the language teacher: a communicative approach, Rowley, Massachusetts: Newbury House, 1984

Service e-Learning Project: State Graduation Online Demo Exam

Ana Mandić
Faculty of Humanities and Social Sciences
Ivana Lučića 3, 10 000 Zagreb, Croatia
ana.mandic@iskon.hr

Ana Bakarić
Faculty of Humanities and Social Sciences
Ivana Lučića 3, 10 000 Zagreb, Croatia
anab246@gmail.com

Nives Mikelić Preradović
Faculty of Humanities and Social Sciences
Ivana Lučića 3, 10 000 Zagreb, Croatia
nmikelic@ffzg.hr

Summary

The service e-learning project described in this paper presents the online demo exam in informatics designed as a preparatory step for the state graduation. It was created in the midst of turbulence caused by the introduction of the state graduation at the end of secondary education in Croatia. With this project we wanted to engage grammar school students in service e-learning and present them with the opportunity to design their own test materials through the collaboration with college students who are experts in the field. We also aimed to gain our own benefit of connecting the theory learned during the study with new practical experiences while at the same time helping the students to achieve at a high level in the state graduation exam. Finally, since its free available online, our project should benefit everyone who wants to test their knowledge in informatics and/or learn something new in an easy and interesting way.

Key words: service e-learning, state graduation, informatics

Introduction

In this paper we present the service e-learning (SL) project, delivered within the course "Service Learning in Information Sciences". The goal of the paper is to present the online exam for the state graduation elective course - informatics, designed to help pupils graduating from grammar school to better prepare for the state graduation. Since this was our first service learning project (actually,

235

our first project at all), we decided to summarize our theoretical knowledge and skills in the field of computer sciences and information sciences and create a demo application that covers the complete information and computer science curriculum of the four-year grammar schools in Croatia. Our ultimate goal was to cover all the curriculum material at all levels, and to address the theory as well as its application in the form of problem solving and design of simple algorithms.

The online demo application is very straightforward: the student gets the results instantly and he can move through the exam at his own pace. The advantage of taking such online exam is that students can get prepared for the real one, reduce stage fright, update and verify their knowledge and discover the areas they are not confident before the real exam.

Basics of Service Learning and its development in Croatia

Service learning (SL) is a teaching method that connects the goals of higher education with the needs of society by providing programs of public service that encourage students to utilize classroom knowledge to improve local communities [1]. It was introduced in the largest faculty of the University of Zagreb (Faculty of Humanities and Social Sciences) in academic year 2006-07 through the series of faculty workshops and through academic courses, with the goal to transform the traditional ex-cathedra teaching style [9]. Goals and requirements for this teaching and learning method were based upon the U.S. experience, gained at the George Washington University. Since then around 50 SL projects in the field of information technology (IT) have been completed and evaluated. Service learning was also introduced in the Faculty of Economy, University of Rijeka in 2008. In 2009 it was added as a regulation of the Croatian National Youth program 2009-2013, approved by the Croatian Government. Also, in the same year the Croatian translation of the "service learning" term ("društveno korisno učenje") offered by one of the authors of this paper became accepted as a common term at the JFDP (Junior Faculty Development Program) Regional Conference. Since the academic year 2009/2010, the Department of Information Sciences at the Faculty of Humanities and Social Sciences in Zagreb offers the stand-alone elective course on service learning (5 ECTS credits). The course "Service Learning in Information Sciences" achieved remarkable student enrolment in a short time and students find this methodology effective, because (according to their course evaluations) it increases their awareness of the world and their personal values and facilitates their engagement and interactivity in the classroom [2, 8].

Since information sciences cover a wide range of topics and due to the fact that information literacy is an important social issue, while the social need for a visual identity (especially in the electronic environment) is constantly growing, information science students truly have a great field for activity where they can meet different interests and apply specific knowledge and skills [3]. Therefore,

236

all the service learning projects in our Department aim at linking the goals of information science studies with IT problems to meet specific community needs.

Comparison with similar research in the USA

Service learning is a teaching strategy currently present in all institutional types and across all fields of study in the United States colleges and universities (see: www.compact.org). As a relatively new teaching strategy, it gained prominence in the U.S. higher education since its emergence in the early 1990s. Its growth in the U.S. is both due to the work of Campus Compact (a national coalition of more than 1,100 college and university presidents supporting student education for responsible citizenship; see: www.compact.org) and Learn and Serve America (a program initiative of the Corporation for National and Community Service - U.S. agency of the federal government; see: http://www.national service.org).

The information science curriculum in the United States applies service learning to facilitate students helping local NGO's on projects related to course topics, such as database design [5] or to connect the students of information science courses with the local schools to provide tutoring in the software applications they are using in class.

Benefits of these service learning projects are reciprocal to both the non-profit organizations that lack monetary resources [7] and to the students learning the problems of the organizations and the potential solutions through technology [11]. Benefits to students also include interacting with real clients and learning interpersonal skills critical to their future [4].

Reflective citizenship and social sensitivity developed on the projects contributes to effective integrated student experiences in the field of information sciences.

State graduation exam

Starting with the school year of 2009/2010, all pupils who complete the fourth grade of grammar schools in Croatia take the state graduation exam (based on the Act on Primary and Secondary Education[1]). Pupils who complete vocational or art schools take the state graduation only if they plan to enroll into polytechnics and universities.

The main aim of the exam is to objectively and impartially test and evaluate the acquired knowledge, skills and abilities of pupils in accordance with the prescribed primary and secondary school curriculum and thus get a comparable assessment of all pupils on the national level.

---

[1] Official Gazette, 87/08

State graduation exam has two parts: mandatory exams in general education subjects: the Croatian language, mathematics and foreign language, and elective exams in one or more optional subjects if these subjects serve as a prerequisite for the enrolment into higher education [6].

In the past graduation system, the results were mutually comparable only on the single school level. In addition, they were susceptible to the subjectivity of the examiners (school teachers). Success on the state graduation is, on the contrary, managed and evaluated by the National Centre for External Evaluation of Education, not by the school teachers. The exam is conducted at the same time, with the same materials and in the same way for all pupils in Croatia, so that the result depends only on the pupil's knowledge and preparedness, not on the secondary school he graduated from. State graduation exam gives pupils a chance to show their real knowledge and skills, since the results are entirely independent from the subjective assessment of any individual teacher. Passing the state graduation exam on the national level, pupils become aware that their success in enrolling into polytechnics or universities depends on the very transparent criteria: their own efforts, hard work and studying. Therefore, the state graduation exam becomes a reliable and objective indicator of pupils' achievement during their primary and secondary education and the valuable motivator for the teachers to increase the quality of elementary and secondary school education. The introduction of state graduation exam promotes and ensures the quality of education, making the selection procedure for continuing education more transparent.

Description of the State Graduation Online Demo in Informatics

State Graduation Online Demo Exam in Informatics[2] consists of 50 multiple-choice questions written in ActionScript3. ActionScript is the object-oriented programming language integral to the Adobe Technology Platform, used to create full-featured Web applications [12]. We have chosen this programming language, because it allowed us to create the exam that can be easily extended and maintained and because it required simple programming techniques, most of which we have already learned during our study. Our aim was to make the elegant code that can easily handle any number of questions.

Pupils usually prefer multiple choice exams, since they find them easier then open questions. They often think that they can recognize the correct answer when they see it, while formulating a correct and complete answer from memory is judged as more difficult. However, being skilled information science students and future teachers, we formulated such options of a multiple choice question so that pupils really need to show the mastery of the subject in order to select the only 100% correct and complete answer.

---

[2] http://cal.ffzg.hr/Ispit_informatika_za_drzavnu_maturu/projekt.html

Multiple choice questions are composed of a question with multiple possible answers, including the correct answer and several incorrect answers (distractors). In order to make the test challenging for the pupils, we did not copy the statements from the textbook but rather used our own words for the question part. Regarding the distractors, we used at least 3 alternatives, plausible and homogeneous distracters, as well as true responses that do not answer the questions [13]. Throughout the test, we avoided responses too close to the correct answer, completely implausible statements and overlapping statements. Finally, we avoided long and complex sentences in the question part, along with trivial responses, negatives, ambiguity and broad generalization.

Although being easy to mark, the questions and answers in the multiple choice exams have to be well formulated. If the pupil interprets the question differently from what we intended, he may give an answer that is correct in a different context, but actually wrong in our interpretation. In this demo exam we applied the following ActionScript3 aspects: variables, control of the playhead of a movie with functions, button event handlers, simple conditionals and text field variables for on-screen display of information. Each question comes with three to four multiple-choice answers. The pupil submits the answer by clicking on the radio button that corresponds to his desired selection. Each selected answer is stored in a variable. Variables with response values are used to grade the user upon completion of the test, with the final scores being displayed on the screen.

Timeline scripts are kept on a single isolated layer, while the series of content layers store various content assets. We define and initialize our main timeline variables before any other scripting occurs. In the next step, we invoke the *stop()* function to keep the user paused on the first frame (where the exam starts). The answer buttons advance the pupil through the exam and keep track of his answers along the way. A message is displayed after each click, showing whether the selection is correct or not. When the playhead lands on our end frame, the calculation script (determining the user's score and displaying it on the screen) is being executed. At the end, a summary screen shows the percentage of the questions answered correctly.

Project evaluation

In order to obtain the objective evaluation of the project, which was initially planned as a support and / or a potential tool for knowledge assessment in informatics, we prepared the evaluation covering all pupils' issues raised during the testing of the applications.

State Graduation Online Demo Exam in Informatics was tested and evaluated by the third grade pupils of Velika Gorica grammar school[3] who will take the state graduation exam at the end of the school year 2011/2012.

---

[3] http://www.gimnazija-velika-gorica.skole.hr

The evaluation was performed as an online survey that aimed to identify the impact of our project on pupils and to discover suggestions that could improve the effectiveness of the exam. The survey consisted of 21 questions that tried to encourage students to critically reflect on the demo exam, but also to reflect on how this way of learning and assessment can further advance the process of preparing for graduation state exam.

Male pupils were slightly overrepresented in our sample: 76.2% of pupils taking the survey were male. The interesting result is that although 33.3% of the pupils in Velika Gorica grammar school never participated in an online exam before, the number of pupils who would like to write the state graduate exams in online environment (47.6%) was only a bit lower than the number of pupils who desire to take the pen-and-paper-based state graduation exams (52.4%).

In regards to the design of the demo exam, 81% of the pupils found it satisfactory, 19% of them thought it was too simplistic, while none of them found it complex.

The overall rating of the demo test was high. We used a five-point rating scale, with "1" being "poor" and "5" being "excellent." The results show that 57.1% of the pupils rated the exam as good, 33.3% gave it a very good mark, while 4.8% of the pupils evaluated it as excellent.

Many of the questions in this survey invited the pupils to state the extent to which they agreed or disagreed with a statement. These responses where then scored as followed: agreement = 3, partial agreement = 2, disagreement = 1. This enabled an average value to be calculated. The Table1 presents the statements divided in two parts: **part A** shows statements related to the pupils' attitudes towards online exams in general, while **part B** covers the statements that refer to the design and usefulness of this specific demo exam in informatics. A high score indicates a high general level of agreement with the statement. Finally, open-ended questions, where pupils were able to offer essay or short answer, sought the subjective pupils' evaluations of this kind of test.

This survey stimulated both us (students) and pupils to critically evaluate the demo project we designed and implemented and raised some new questions regarding its functionality and possible future application.

All of the pupils' suggestions will be used in the further development of the project, and thus will naturally increase the benefit for all future users who will take this exam in order to prepare for the state graduation exam.

Number of points that each pupil has earned in this demo exam was used to prepare the statistical results for the school teachers, but also to assess the level of knowledge needed to successfully complete the online exam and to modify the exam in terms of facilitating questions or by introducing more complex tasks.

Table 1: Online survey results

Question/statement	Average score (3 for agreement, 1 for disagreement)
PART A	
Before taking this demo exam, I had positive attitude towards online exams.	2.38
After completing this demo exam, I have positive attitude towards online exams.	2.38
I feel more comfortable with taking test on a computer instead of pen and paper.	2.24
In my further education, I want ALL written exams to be conducted online.	2.09
In my further education, I want for SOME written exams to be conducted online.	1.05
PART B	
State Graduation Online Demo in Informatics was useful and interesting.	2.52
The immediate feedback in online test was comprehensible and useful.	2.38
The immediate feedback in online test provided me with the opportunity to learn from wrong answers	2.48
This demo exam would be more appealing and more usable if it had time constraints.	1.90
In order to be effective, this demo exam should also contain another type of questions, such as fill-in-the-blank or short answer	1.95
The type of questions in the demo exam is not important, as long as it covers the entire content in a subject of informatics for the 4-year grammar schools	2.29
Clicking the "Check" button next to each answer in the test is unnecessary and time consuming	2.10
Clicking the "Check" button next to each answer in the test is useful and efficient.	2.19

Although the project has received overall positive feedback and response, opinions on the online demo exam are divided. Our assumption is that the reason for this is the fact that the majority of pupils never had a chance to take exams via computer and get immediate feedback on the accuracy of their answers. In any case, the pupils expressed their motivation and positive attitude towards online tests and learning. They have recognized the usefulness of knowledge assessment through the online examination, since they received feedback on accuracy, but also the opportunity to learn from the incorrect answers, which was particularly stressed as an advantage of this application.

Furthermore, pupils assessed the extent of the material covered as more important than the interface design but they also believe that the multimedia component would strengthen a pupil's interest and appreciation of the both subject material and this type of examination.

Finally, the type of questions and time constraints of the exam were not considered too important, although most of the pupils shared the opinion that the exam would be more appealing if the response-time constraints were introduced. They believe that these constraints would encourage pupils to provide answers to more questions in the same amount of time.

Regarding the service e-learning component, this project contributed to the pupils' preparedness for the state graduation exam in elective subject - informatics and offered them an insight into new technology and new ways of knowledge acquiring and its evaluation. The questions in the demo exam cover the entire content of the subject of Informatics for the 4-year grammar schools, while the simple but interactive online application enables pupils to take the exam and test their knowledge anytime and anywhere, getting the immediate feedback.

Conclusion and future work

Amid all the turmoil surrounding the introduction of the state graduation exam in Croatia, we came up with the idea to unify our knowledge in e-learning, programming and theory of information sciences and to turn it into hands-on experience that will satisfy at least one community need. This type of the project can be developed for any school subject, with different types of questions. Apart from the objective type questions, such as true-false, matching, multiple choice, fill-in-the-blank, numeric response or some combination of the above mentioned, online exam in some subjects, such as the Croatian language, can also consist of subjective type of questions, such as short answer or essay.

One of the main benefits of the online exam is the immediate availability of the results. Physical presence at a given location is not necessary for both the pupil and the teacher since it's published on the Internet and available anywhere in the world. It is economical (there is no need for printing and copying materials), it can be designed with the use of multimedia, which increases the attention and the motivation of the user and exam can be given to pupils as extensive training even for subjects that are hardly to have the actual examination conducted in an online environment. The main disadvantage is its high dependence on the connection speed: for instance, dial-up would limit the use of graphics or media files.

This project is important for all elementary and secondary school teachers searching for ways to incorporate the e-learning into classrooms on both a short-term and long-term basis, to address individual learning traits and spark learning excitement. One of the ways to do this is by providing pupils with online tests to prepare for hand-written tests [10], but also to adapt to the new conditions in the higher education institutions where the use of online exams, online courseware, tutorials and modular internet-based courses is rising dramatically.

Regarding our benefit in this service learning project, we gained a substantial understanding on how teaching informatics in school looks like. Working with

242

pupils, we gained the ability to adapt to their feedback and learned the importance of listening to their voice and to be opened to the ideas of others. We realized that we have to work with both school teachers and pupils to select strategies to increase popularity of online exams among pupils. Therefore, we believe our project is only the first step towards the collaborative design and implementation of the new educational environment in which pupils can use and further develop their IT skills in a productive and satisfying way.

Finally, apart from the real-world experience and strengthened portfolio, this project gave us the opportunity to make real decisions, to engage our knowledge and skills and to use them to do something that has impact on others beyond ourselves. We gained insight into our strengths and weaknesses and learned more about each other as well.

References

Bringle, R., Hatcher, J. Implementing Service Learning in Higher Education, Journal of Higher Education, Vol.67, No.2, March/April, 1996.

Eyler, J., Giles, D., Stenson, C., Gray, C.: What we know about the effects of service-learning on college students, Faculty, Institutions and Communities, 2001.

Eyler J, Dwight E. G. Where's the Learning in Service Learning? Jossey-Bass Publishers, 1999.

Guthrie, R.A. & Navarrete, C.J. Service-Learning Impact on Information Systems Students in a Web Development Course // Proceedings of the Information Systems Education Conference (ISECON). San Diego, CA, 2003.

Hoxmeier, J. & Lenk, M.M. Service Learning in Information Systems Courses: Community Projects That Make a Difference // Journal of Information Systems Education (14), 2003, p. 92.

Janjic, Z. et al. (eds). Drzavna matura 2010./2011. i prijave za upis na studijske programe. // Ministarstvo znanosti, obrazovanja i sporta, Agencija za znanost i visoko obrazovanje, Nacionalni centar za vanjsko vrednovanje obrazovanja, Hrvatska akademska i istraživačka mreža CARNet, Zagreb, 2010.

Lazar, J. & Norcio, A. Service-Research: Community Partnerships for Research and Training // Journal of Informatics Education and Research (2)3, 2000, pp. 21-25.

Mikelic Preradovic, Nives; Kisicek, Sanja; Boras, Damir. Evaluation of Service learning in ICT curriculum // Proceedings of the 2nd Paris International Conference on Education, Economy and Society, Vol. 3: Analytrics / Tchibozo, G. (ed). Strasbourg (France): Association inscrite au Registre des Associations du Tribunal d'Instance de Strasbourg, Volume 85, Folio n° 1, 2010. pp. 55-66.

Mikelić, Nives; Boras, Damir. Service learning: can our students learn how to become a successful student? // Proceedings of the 28th International Conference on Information Technology Interfaces / Budin, Leo; Lužar-Stiffler, Vesna ; Bekić, Zoran ; Hljuz Dobrić, Vesna (eds). Zagreb: SRCE, 2006, pp. 651-657.

Pont, R. Do Online Quizzes help Students prepare better for Hand-Written Tests. // California State University Northridge, 2009. http://www.csun.edu/~rjp79535/coursework/697/research project.pdf (Accessed: July, 5th. 2011.)

Saulnier, B.M. Service Learning in Information Systems: Significant Learning for Tomorrow's Computer Professionals // Proceedings of the Information Systems Education Conference (ISECON). Newport, RI, 2004, p. 9.

Shupe, R. Learning Flash CS4 Professional. O'Reilly, 2009.

Ward, A.W., & Murray-Ward, M. Assessment in the Classroom. Belmont, CA: Wadsworth Publishing Co., 1999.

Erasmus Intensive Programme: Information and Communication Technology in Supporting the Educational Process

Mihaela Banek Zorica
University of Zagreb, Faculty of Humanities and Social Sciences,
Department of Information Sciences
I. Lučića 3, Zagreb, Croatia
mbanek@ffzg.hr

Lucia Amante
Universidade Aberta, Education and Distance Learning Department
Campus do Taguspark, Edifício Inovação 1, Av. Dr. Jacques Delors, 2740-122,
Porto Salvo - Oeiras, Portugal
lamante@univ-ab.pt

Gloria Bastos
Universidade Aberta, Education and Distance Learning Department
Campus Taguspark, Edifício Inovação 1, Av. Dr. Jacques Delors, 2740-122,
Porto Salvo - Oeiras, Portugal
gloria@univ-ab.pt

Serap Kurbanoglu
Hacettepe Üniversitesi, Faculty of Letters,
Department of Information Management
Beytepe, 06800 Ankara, Turkey
serap@hacettepe.edu.tr

Sonja Špiranec
University of Zagreb, Faculty of Humanites and Social Sciences,
Department of Information Sciences
I. Lučića 3, Zagreb, Croatia
sspiran@ffzg.hr

Yasar Tonta
Hacettepe Üniversitesi, Faculty of Letters, Department of Information
Management
Beytepe, 06800 Ankara, Turkey
tonta@hacettepe.edu.tr

Vlasta Zabukovec
Univerza v Ljubljani, Filozofska fakulteta, Oddelek za bibliotekarstvo,
informacijsko znanost in knjigarstvo
Aškerčeva 2, 1000 Ljubljana Slovenija
Vlasta.Zabukovec@ff.uni-lj.si

245

Summary

This paper describes the Erasmus IP course "Information and communication technology in supporting the educational process (ICT-EE)" organized by the Department of Information Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, that has taken place from September 10th to September 23rd 2011. In this IP, participants from four universities with a total of 10 teaching staff were involved in the project: University of Zagreb, Croatia (http://www.unizg.hr), Universidade Aberta, Portugal (http://www.uab.pt), Univerza v Ljubljani, Slovenia (http://www.ff.uni-lj.si) and Hacettepe Üniversitesi, Turkey (http://www.hacettepe.edu.tr/english/index.php). The wide range of expertise of the teaching staff gathered around this IP presents a strong multidisciplinary approach whose focus was on strengthening international collaboration, offering different insight to the issues of e-learning as well as bringing together specific and unique expertise that can be applied in any given field of education. The main objective of this course was to give students insight into the problem of teaching and learning in electronic educational environment.

Key words: Erasmus IP, higher education, library, e-learning, information literacy

Introduction

Information and communication technology has become essential in teaching and learning process in higher education and research. As the quantity of the available digital content is increasing, students as future employees and users of information institutions need to acquire knowledge which will give them competencies for understanding and development of new user services based on the digital content.

The Erasmus Intensive program[1]: Information and communication technology in supporting the educational process (ICT-EE) as a mode of short program study was seen as a suitable mode of knowledge transfer between students from different partner countries. Program based on the issue of successful utilization of modern information technologies in educational process was seen as a necessary extension to the existing curricula at the partner universities. The main objective of the intensive programme was that of creating multidisciplinary and multilateral networks between teachers and students from partner EU and non-EU higher education institutions which, due to their origins, educational missions and objectives, offer different approaches and context to the topics covered. The wide objective of the IP was to bring together professors and students

---

[1] "An Intensive Programme (IP) is a short programme of study which brings together students and teaching staff from higher education institutions of at least three participating countries. It can last from 10 continuous full days to 6 weeks of subject related work." ERASMUS Intensive programmes http://ec.europa.eu/education/erasmus/doc900_en.htm

from EU and non-EU countries to increase their knowledge in educational technology and support. The wide range of expertise of the teaching staff gathered around this IP presents a strong multidisciplinary approach whose focus is on strengthening international collaboration, offering different insight to the issues of e-learning as well as bringing together specific and unique expertise that can be applied in any given field of education. The goal of activities planned was to promote training and research among body of students and encourage international collaboration by which the full potential for European identity, citizenship and employability of adult learners with mobility along with the opportunities for their personal growth will be encouraged. It tried to promote teamwork and citizenship, engaged problem-solving in the specific context (rather than generalized or abstract concepts from a textbook) and gave participants the ability to make connections across the disciplines.

About the ICTEE – program overview

The course was based on a set of theoretical and practical tools related to cross-disciplinary education and educational procedures, tools and supporting elements facilitating thus the application of students' theoretical and conceptual knowledge on practical examples that can often be found in the educational process. The main objective of this course was to give students insight into the problem of teaching and learning in electronic educational environment. The goal was to provide students with necessary knowledge and skills which will enable them to explore and connect theory with real life problems and challenges by utilizing hands on experience and learning.

Added value of the Erasmus Intensive program (IP) was in unique mix of different aspects of the education in electronic environment covering all the supporting issues like libraries, intellectual property, service learning and technology usage. The program was organized in general direction by which learning outcomes could become applicable in any field offering an interdisciplinary approach that could be utilized in any educational context.

In this IP, participants from four universities with a total of 10 teaching staff were involved in the project. As Universidade Aberta is specialised in distant learning and therefore was chosen to provide input in the aspects of the e-learning which they have most experience with. Specifically, topics covered were: Assessment in e-learning; and Libraries and e-learning. Furthermore, Univerza v Ljubljani, Filozofska fakulteta, Oddelek za bibliotekarstvo, informacijsko znanost in knjigarstvo puts emphasis on psychological and pedagogical role and tasks of librarians, therefore their input was seen in the topic: Psychology and LIS environments. For successful education in the digital age there is a growing need for recognition and understanding of intellectual property rights in the digital age; and development of web-based information literacy packages which were the topics offered by the Hacettepe Üniversitesi. Information literacy has proven to be the key component in successful learning and

acting in the digital age; therefore the Univerisity of Zagreb offered the following topics: Quality management in e-learning, E-portfolio, Service e-learning and Speaking in the electronic environment.

The primary target group were students of Master's degree level, although bachelor's or PhD degree students were also accepted. Selection process was done at each partner institution where the best students were chosen. During the application process of the IP all partners agreed on accrediting student participants with the assigned 5 ECTS. The practical exercises have been organized in international subgroups, making it possible for the participants to interact with fellow students from different countries and cultural backgrounds. The IP course activities have consisted of lectures followed by discussions, seminars, working groups, and subject-related excursions. Educational activities were complemented by cultural activities related to the host country heritage as well as IP subject matter. A special focus was on the interactivity, therefore various types of sessions have been offered, involving student in both theoretical and practical aspects of the ICT in educational environment. The technology used in the IP presents a mixture of various types of technology that could be used in education, encompassing different aspects of educational process. Therefore, students have been able to gather insight into the technology currently used at the higher education institutions such as: various web 2.0 tools (Gloggster, Voicecaster, Xtranormal, ScoopIT etc), VOIP, Virtual worlds (Second life), presentational tools (Prezi), Course management systems (Moodle) etc.

Students have produced live posters in their group-work. These have been presented using one of the new media tools which served as discussion points in the closing session at the end of the IP. The virtual learning environment of the IP offered through the Moodle open-source learning platform whose FHSS implementation is called Omega. This enabled students to access learning materials before, during, and after the IP. Selected papers presented by students at the IP were made available to the general public on the project website (http://infoz. ffzg.hr/erasmusip/).

Students were offered to actively participate in lectures presented by professor(s) and by presenting their point of view and examples about topics presented in lectures and in discussions. They will also participate actively in short exercises corresponding to the topics presented in lectures. Direct students' participation in transfer of knowledge will enable higher quality of teaching. By participating in solving real-life problems during lectures, discussions and exercises, students will be able to evaluate the level of knowledge they acquired at the end of the lectures and to compare this level with the objectives set for the topic. At the end of lectures, students were given an evaluation survey for the course which they evaluated as of high quality and necessary in their future work environment still not crucial in their job search.

Project outcomes

One of the results of the IP is strengthened cooperation between educational European institutions in performing ICT supporting lifelong learning. Program has tried to ensure transfer of ideas and different methodology as well as evaluation criteria of participating countries. New teaching methods for lifelong learning enabled participants to learn how to implement tools and methods in their own personal growth but also in their future professional work either as teachers or professionals involved in lifelong learning or learning organizations.

Learning outcomes of the IP to address the following:

- to know key components of modern education that consists of a well balanced use of new emerging technologies;
- to understand the electronic educational environments'
  - methodological approaches
  - supporting elements and
  - quality management issues
- get insight in key issues and problems.

The outcome was creation of ICT-based content, i.e., producing teaching materials, student reports/theses, multimedia products and so on. Planned activities enabled participants to learn together with fellow students from other European higher education institutions which help them in development of intercultural competence as well as improvement of their foreign language proficiency. We hope that multidisciplinary approach to knowledge development and usage of ICT technologies contributed to transparency and universal understanding of public knowledge, educational paradigm, social communication, media and information literacy, etc. in a global world.

Student work has been evaluated by the teaching staff. By using new media such as Twitter, students were processing and disseminating what they have learned among themselves but also with the general public by creating stream #ICTEE11. Furthermore, close connection among students that has been created with this IP, is seen in their connecting through social network (Facebook) and creation of the group dedicated to and consisting of participants of the course.

Conclusion

The main task of the ERASMUS IP ICT-EE was to strengthen cooperation between educational European institutions in researching and utilizing ICT and in supporting lifelong learning. The wide range of expertise of the teaching staff gathered around this IP presents a strong multidisciplinary approach whose focus is on strengthening international collaboration, offering different insight to the issues of e-learning as well as bringing together specific and unique expertise that can be applied in any given field of education. The main objective of this course was to give students insight into the field of education in electronic

environment. The IP course activities were a combination of lectures followed by discussions, seminars, working groups, and subject-related and cultural heritage excursions.

References

Crowther, J. ; Sutherland, P. *Lifelong Learning: Concepts and Contexts*. New York: Routledge, 2008.

Erasmus Intensive programmes http://ec.europa.eu/education/erasmus/doc900_en.htm (25.06.2011)

ERASMUS ICT-EE http://infoz.ffzg.hr/erasmusip/ (25.06.2011)

Garrison, R.; Anderson, T. *E-learning in the 21st Century*. London: Routledge Falmer, 2003

Morgan-Klein, B.; Osborne, M. *The Concepts and Practice of the Lifelong Learning.* Osborne, M, Sankey, K and Wilson, B. (Eds). London: Routledge, 2007.

Pereira, A.; Quintas-Mendes, A.; Morgado, L.; Bidarra, J. *Universidade Aberta's Pedagogical Model for Distance Education. A University for the Future.* Lisboa: Universidade Aberta, 2008.

SOCIAL COMMUNICATION

Understanding Social Media Acceptance and Use in the Context of Technology Generations and Life-Based Design

Jaana Leikas
VTT Technical Research Centre of Finland
Tekniikankatu 1, P.O.Box 1300, FI-33101 Tampere, Finland
E-mail: jaana.leikas@vtt.fi

Mari Ylikauppila
VTT Technical Research Centre of Finland
Tekniikankatu 1, P.O.Box 1300, FI-33101 Tampere, Finland
E-mail: mari.ylikauppila@vtt.fi

Jussi Jokinen
University of Jyväskylä
Agora Center, P.O.Box 35, FI-40014 University of Jyväskylä, Finland
E-mail: jussi.jokinen@jyu.fi

Rebekah Rousi
University of Jyväskylä
Agora Center, P.O.Box 35, FI-40014 University of Jyväskylä, Finland
E-mail: rebekah.rousi@jyu.fi

Pertti Saariluoma
University of Jyväskylä
Agora Center, P.O.Box 35, FI-40014 University of Jyväskylä, Finland
E-mail: ps@jyu.fi

Summary

People's relation to technology, and especially to ICT, is changing. A good example of this is the existing and emerging social media which allows people to build new types of networks with each other. As it is a new phenomenon which has become an important part of life to many people, it is logical to ask if differences can be found between people of various ages in their relation to social media. Here, the differences and similarities between technology generations are examined. The study begins with the idea of a cohort learning and adopting technology, and gives insights about the relevance of a technology generation paradigm in the context of life-based design.

Key words: design, life-based design, social media, technology generations, human-technology interaction

253

Introduction

Modern ICT technology, mobile and fixed, allows people to use the same device for practically an unlimited number of purposes. The possible purposes people have for using technologies are becoming a vital topic in the science of human-technology interaction (HTI) design and design discourse. In technology development this means that finding new ways of investigating the relationship between people and technologies have become central. As it is possible to find numerous ways of carrying out the design process with users it is time now to ask what the rational forms of HTI design can be.

Life-based design is a design paradigm with the key idea of investigating the HTI design process in the concepts pertaining to research into human life (Leikas and Saariluoma 2008; Leikas 2009; Saariluoma and Leikas 2010). Here, the sociology, psychology and biology of human life form the basic conceptualization, methodology and factual basis for design (Leikas 2009; Saariluoma and Leikas 2010). In our research we apply life-based design methodology to investigate differences in social media use between generations and in different stages of life.

Technical devices and services motivate people by the added value they bring to everyday life. The added value of technology can be found in the practical outcome of technology use (Cockton 2006; Melenhorst 2006; Bohn et al. 2004). In other words, the end results prove that an outcome has been reached with the help of technology, which could not have been achieved without the tool in question. This is why it is valuable in designing new technological ideas to have a clear idea about the expectations and goals of the potential users. Here, understanding the demands harboured by people of different ages is significant, as age impacts the way in which people utilise technology and more importantly influences the goals of usage in everyday life (Leikas and Saariluoma 2008; Morris, Venkatesh and Ackerman 2005). Accordingly, the connection between age and technology use is an important problem in life-based design (Leikas 2009).

A concrete challenge today in HTI discussion is to understand the differences in usage of social media by people of different ages. Services such as Facebook, Twitter and Orkut have promptly found a place in human everyday life. These services create possibilities for social contacts, which can promote psychological well-being and enhance the quality of life (Gollwitzer, Delius and Oettingen 2000; Jones and Fox 2009). Today, we still have relatively little information about the use of these services though the number of users can be measured in hundreds of millions (Lievrouw and Livingstone 2002; Boyd and Ellison 2007). For life-based design research, this aspect of the information society is central.

The notion of technology generation has become valuable in research into the effects of age in ICT-use. Technology generations reflect the historical timing of computing innovations and their diffusion into productive and cultural spheres, linked with the time period in which a cohort comes of age (McMullin,

254

Duerden Comeau and Jovic 2007). One reason for understanding this concept in HTI design is that people learn to use technologies at a certain age, and this understanding of how to use technologies (present and future) is built on the kind of knowledge that is typical for that cohort (Docampo Rama 2001). For example, Sackmann and Weymann (1994) introduce different generations based on the experience of technology usage available in the formative period.

Studies have outlined the different technological eras (Lim 2010) generally divided into the 'mechanical' (M) era (born before 1930), the 'electro-mechanical' (EM) era (born 1930 – 1960), and the 'digital-software' (DS) era (born after 1960). Lewis, Langdon and Clackson (2007) have categorized the time era during which one was born in relation to interface technology. These categories are 'the electro-mechanical' era (born pre-1928), 'sees the remote control' era (1928-1964), 'dominated by displays' era (1964-1990), and the era where post 1990 layered menu systems are generally prevalent and popular.

Similar to what has been said above, four major interaction styles of consumer products have been identified by Docampo Rama (2001). These are the mechanical style (1930/1940), the electro-mechanical style (available 1930-1980), the display style (available 1980-1990), and the menu style (implemented 1990 ->). Docampo Rama (2001) goes on to introduce three technology generations, which are: the electro-mechanical generation (born 1930-1960, + before 1930); the display generation (born 1960-1970); and the menu generation (born after 1970). McMullin, Duerden Comeau and Jovic (2007) have studied the phenomenon from an ICT technology point of view and have discovered five technology generations. These are: the pre-ATARI generation (born prior to 1955), who came of age before computing technology had widespread cultural appeal or was widely used; the ATARI generation (born 1955-1963) to whom the ATARI home video games became popular and the first PCs were introduced in the workplace; the console generation (born 1964-1973) who used Commodore64, TRS-80, Tetris, Apple MacIntosh and Windows 3.0, and who had great opportunity to use new computer technology at home; the Windows generation (born 1974-1978) who used Microsoft, Windows 97/98, Excel, Adobe pdf, Email, SimCity, Doom and witnessed the launch of the Internet in the mid-1990s, although it was not immediately or widely embraced; and finally the Internet generation (born after 1978) who are familiar with the Internet, Yahoo, Google, Instant Messaging (MSN), Windows XP and iPods.

Yet another way to categorize users according to their age is to classify them as Boomers (born 1946 to 1964), Generation X (born 1965 to 1976) and Generation Y (born 1977 and younger) (LexisNexis 2008; Tapscott 2009; Stutzman 2010). The introduction of television defined the technological tone of the baby boom. Generation X, the "baby bust", is a much smaller cohort. Though its members are quite educated and regularly use advanced communication technology, they did not grow up with computers, and many feel somewhat excluded from the central cultural debate. Tapscott (2009) also talks about the Net

Generation (born 1977 to 1997), which is again a larger group, a kind of "echo" of the baby boom. This cohort has been around computers since before they could speak. For them, technology is a necessity, yet it is invisible. They cannot imagine living without technology, and their continual connection to others worldwide has produced the first truly global generation.

Social media is emerging technology which has become vital for our future. Therefore, it is necessary to understand the possibilities it opens and the demands it sets for users of different ages. Stutzman (2010) argues that if we examined the communication patterns of people, we would find that cohort-level structures guide a good deal of the use of social technology. Differences can indeed be seen in technology and social media usage between generations. A Pew Internet report (Lenhart et al. 2010) reveals that 93% of American teens ages 12-17 go online, as do 93% of young adults ages 18-29. Three quarters (74%) of all adults ages 18 and older go online. Over the past ten years, teens and young adults have consistently been the two groups most likely to go online, even as the internet population has grown and with larger documented increases in certain age cohorts (e.g. adults 65 and older).

The survey of LexisNexis (2008), with the total sample size of 700 American legal and white collar professionals, reveals differences in technology acceptance among generations. According to the survey, two-thirds of all Boomers agree that Personal Digital Assistants (PDAs) and mobile phones contribute to a decline in proper workplace etiquette, and believe the use of a laptop during in-person meetings is distracting, whereas less than half of Gen Y workers agree on this. Only 17% of Boomers believe using laptops or PDAs during in-person meetings is efficient, while more than one third of Gen Y do. Only 28% of Boomers think blogging about work-related issues is acceptable, while forty percent of Gen Y workers do.

According to LexisNexis (2008), Gen Y workers (ages 31 and younger) generally do not see a problem accessing personal web sites from work - like Facebook and blogs, but only 14% of Boomers access a social network from work. In fact, Gen Y spends a lot more of their day online accessing social networks, news sites, blogs, forums, and multimedia sharing sites than the Boomers. Also, Gen Y switches back and forth between applications far more than the Boomers do.

This brief introduction to technology generations makes it understandable that in developing social media, we have to have a solid understanding of the problems of technology generations. For this reason, we searched for the differences between technology generations in the factors influencing different usage patterns.

Empirical study of usage patterns and technology generations

In our study, we explored the usage of social media within different technology generations. Doing this we have tried to isolate whether or not differences exist between the technology generations, both in terms of generational categories and in terms of age, and what could be possible implications for design. Our study has focused on the examinations of different usage patterns of social media. In HTI, usage patterns have generally been examined in order to sort people as technology users (Prendergast and Roberts 2008; Blyth and Roberts 2005; Shove, Watson and Ingram 2005; Shove and Pantzer 2005; Silverstone 1992). We explored usage patterns of social media between different technology generations and examined whether the generational formation could give insight into analysing different life settings in the context of social relationships. The next section elaborates on the theoretical and methodological bases upon which we have organised the study.

Method

Six focus group sessions were held in total, discussing different life settings and overall opinions about social media. The key rationale here is that the participants are the informants of their everyday life, and the themes given by the moderator provide a conceptual basis through which participants can imagine themselves as users of different social media services. In a focus group it is possible to uncover problems in the usage situations of existing services and identify space for improvement or modifications.

First the participants were asked to undertake a questionnaire concerning how they consume and perceive online services and especially social media. Once the questionnaires were completed, the focus group discussions were structured around the topical sections of the questionnaire. These sections were: usability of existing products, technology generations, life settings (forms of life) and social relationships, as well as privacy and trust. In the focus groups we considered and discussed what social media really is, how social relationships can be enhanced and e.g., how loneliness can be recognised and reduced via the help of technology, and the types of things people do not want to accept through technology. In other words, the groups were used as forums to discuss what technology had to offer, and what it could not offer in relation to different situations in life. The prospect of a hypothetical shopping mall for different situations in life was used to inspire thoughts and ideas to address this area.

The questionnaires were analysed quantitatively, and the focus group audio recordings were analysed and coded via content analysis. The results were analysed both quantitatively (questionnaire) and qualitatively (audio recordings). Content analysis was used to decipher key themes mentioned amongst the participants.

257

Participants – samples and generations

In total, 31 Finnish citizens and 14 Australian citizens participated in the group discussions. In total there were 13 male and 32 female participants: in the Finnish group there were 23 women and 8 men; in the Australian group there were 9 women and 5 men. The participants were selected according to the following generational categories, which were defined based of the earlier studies of technology generations introduced earlier. The categories were seen to best match the historical development of technology in Finland. These are:

- Electro/Mechanical Generation (born before 1965),
- Layered Interface (the Microsoft) Generation (1965-1980), and
- Games Generation (1981-1992).

The groups were distributed as follows: born before 1965 age group – 22 participants: 18 Finns (5 men and 13 women) and 4 Australians (1 man and 3 women); 1965-1980 – 9 participants: 6 Finns (2 men and 4 women) and 3 Australians (3 men and no women; and 1981-1992 – 14 participants: 7 Finns (1 man and 6 women) and 7 Australians (1 man and 6 women).

Results

In this paper, we report the results of the questionnaire study carried out in the context of the focus groups sessions. The participants were divided into three technological generation groups. But do these predetermined groups represent the reality? In order to test the validity of the three generations outlined above, a gottman sum variable was constructed. The variable represents the number of teenage technological experiences. The participants were asked to mark down how old they were when they had their first experience with the internet, email, social media, mobile phone etc. If the participant was under 20 years of age during the first encounter with the technology, the sum variable got an increase of one unit.

The sum variable had a minimum of zero and a maximum of six teenage technological first encounters. In the data, the mean of this variable was 2,11 (standard deviation 2,48). Half of the participants had no teenage experiences with information technology, and one third had more than 4. We used four or more technological teenage experiences as a cut line in order to determine if the participant was part of the newest technology generation. A cross tabulation confirmed that the technology generations explicated above conform to the empirically constructed test variable (Pearson chi square test, $\chi^2=37,8$, $p<,001$). Those who had four or more teenage IT first encounters were exclusively classified into the third generation. This is understandable, as most of the technologies queried did not exist before 1980s.

Our second area of interest was computer skills. The participants born before 1965 evaluated their own computer skills as weaker (mean 2,50, standard devi-

ation ,802, on a scale from 1 to 5) in comparison to those born from 1965 to 1980 (mean 3,22, standard deviation ,833) who had evaluated their own skills the highest out of all the generation groups. Finally, those born from 1981 to 1992 (mean 2,64 standard deviation ,650) evaluated their own skills stronger than the eldest participants yet still not as strongly as the middle generation. This difference in means was revealed statistically different by one way anova ($F(2,44)=3,212$, p=,05).

The frequency of technological activity was queried using two variables, one for activity at work and one for activity during free time. Not surprisingly, the usage of technology was very frequent among the participants both at work and during free time. Technological activeness did not correlate with technological skills, but for both nationalities activity at work correlated with activity during free-time (r=,395, p<,05). This finding can be elaborated by comparing the correlation between the technology generations. Spearman correlation between the groups revealed, that only for the third and youngest technology generation, the correlation between technology activity at work and during free-time is statistically significant (r=,690, p<,01).

The use of social media was frequent among the participants. Fifty-one percent of those who used social media used it daily and only 9% used it only once a week. This frequency was not associated with any technology generation: all age groups were equally active in their usage of social media. The participants use social media at home as well as at work or at school, except for the oldest generation, which understandably does not utilise social media at work. All technology generations responded with high technology activity (mean around 4,5, i.e. "almost daily") both at work and at home.

The next area of interest was trust in social media. This was examined by using different Likert-scale questions, like willingness to reveal personal information such as name or date of birth. These variables were combined into a sum variable (α=,892), which on a scale of 1 to 5 describes the overall trust in social media (five being great trust). The mean for trust in social media was 2,36 (standard deviation 1,06), and one way Anova revealed statistically different means between the technology generations ($F(2,40)=5,398$, p<,05). Not surprisingly, the increase in age predicts a considerable decrease in trust towards social media (r=-526, p<,001).

In accordance with Leonardi et al. (2008), there seems to be general hesitation to reveal too much within social media. Most of the born-before 1965 participants revealed within the discussion, an unwillingness to accept unknown Facebook users as friends. Likewise, there was general consensus among all participants, that friends of friends would not be accepted as friends.

Discussion

We have studied the nature of social media and investigated some of its main features in relation to technology generations. The overall picture is clear. Different generations use social media on a general level rather similarly, but important differences can also be found. Understandably, the basic use of social media is rather similar between different generations. However, young people use special functionalities of social media, such as different discussion groups, more than older users. Older people in turn have much less trust in social media than young people. Possibly, these two phenomena are linked to each other, as being involved with groups usually presupposes trust.

The similarities and differences of social media usage can be considered in the context of life-based design. Life-based design looks at the contexts, events, values, motivations and technologies that shape the cohort over time. For example, life settings of people may differ according to the combination of a number of factors, such as age, family and marital status, social status, profession, health issues, education, gender, skills, and, as introduced here, technology generations. These factors ultimately impact everyday needs related to communication and companionship. These in turn influence the ways in which people experience and evaluate available products and services, and what kind of added value technology could offer them.

The ground idea of life-based design is that it is necessary to consider technology always in the context of human life. The design of technology should begin and end with life (Leikas and Saariluoma 2008; Leikas 2009; Saariluoma and Leikas 2010), and design thinking should be grounded on the concepts of human life science. This means sociology and corresponding sciences such as ethnography, organizational research and management, psychology and its close allies such as philosophy of mind, education, and ergonomics, and finally, biology as well as other related areas of research such as medicine, neuroscience or physiology and anatomy. This list can be added to multidisciplinary areas of research such as cognitive science, gerontechnology, occupational therapy, design science and art design. The key unifying argument is that the problems of human-technology interaction design should be conceptualized and argumentatively supported in concepts and theories developed for the analysis of human life.

Technology generations open an interesting multidisciplinary discourse in life-based design. This problem begins with the idea of a cohort for learning and adopting technology. This means a cohort, which is in part a life history that has been characterized by certain technologies adopted earlier in life. A cohort is a sociologically relevant issue, whereas learning and its transfer through life are important psychological issues. Age groups are also biologically and mentally different in many respects. This is why the problem of technology generations is so vital as it offers a tool to understand people's technological behaviour and

the acceptance of new technologies by studying usage patterns of different technology generations.

Our first study to investigate the role of technology generation-related problems in life-based design illustrated two important features for designers. Firstly, in technological usage there are many standard features, which in turn means that skilled and interested users do not differ much from between each other. Hence, generations do not differ significantly from one another in any essential manner either. Therefore, designers normally look for optimal and working solutions for all 'occasional' users. However, knowing the areas in which there are no significant differences between technology generations can sometimes be vital.

Secondly, there are issues such as trust and privacy in which generations do differ. This means that the designers must understand the requirements set for their work by the nature and background of the specific difference. This is a presupposition for successful design. The designers must understand the logic which differentiates the different generations, and be able to create individually modifiable technologies or find means to set aside such emotional issues as trust and privacy problems.

Future work

When thinking of technology such as social media as a means for enhancing social relationships, much work is yet to be done in terms of addressing the working age- beyond working age divide. What would be e.g., the drive behind encouraging older adults to use these platforms as a means for branching social contacts?

We shall tackle this issue further in our future study of technology generations and forms of life. The study will collect data from different European countries and Australia about technology and life settings and hopefully reveal more about this interesting problem in life-based design.

Acknowledgements

References

Blyth, S., and Roberts, S. Re-thinking market research. Paper presented at the MRS Conference 2005, London, 2005

Bohn, J., Coroama, V., Langheinrich, M., Mattern, F. and Rohs, M. Living in a World of Smart Everyday Objects-Social, Economic, and Ethical Implications. Human and Ecological Risk Assessment, 10 (5), 2004 October, 763-785.

Boyd, D., and Ellison, N. Social Network Sites: Definition, History, and Scholarship. Journal of Computer Mediated Communication, Vol. 13, No. 1 (2007, October), 210-230Cockton, G. Designing worth is worth designing. In Proceedings of the 4th Nordic conference on Human-computer interaction: Changing roles, 14-18 October 2006, Oslo, 2006, 165-174

Docampo Rama, M. Technology Generations - Handling Complex User Interfaces. Eindhoven: University of Eindhoven, 2001

Gollwitzer, P.M., Delius, J.D. and Oettingen, G. Motivation. In Pawlik, K. & Rosenzweig, M.R. (Eds.), International handbook of psychology. London: Sage, 2000

Jones, S. and Fox, S. Generations online. Pew Internet report 2009. http://www.pewinternet. org/Reports/2009/Generations-Online-in-2009.aspx (18.6.11)

Leikas, J. and Saariluoma, P. 'Worth' and mental contents in designing for ageing citizens' form of life. Gerontechnology, Vol. 7, No. 3 (2008), 305–318

Leikas, J. Life-Based Design - A holistic approach to designing human-technology interaction. VTT Publications 726. Helsinki: Edita Prima Oy, 2009 http://www.vtt.fi/inf/pdf/publications/ 2009/P726.pdf

Lenhart, A., Purcell, K., Smith, A. and Zickuhr, K. Social media and young adults. Pew Internet Report., Feb 3 2010. http://pewinternet.org/Reports/2010/Social-Media-and-Young-Adults/ Summary-of-Findings.aspx (18.6.2011)

Leonardi, C., Mennecozzi, C., Not, E., Pianesi, F. and Zancanaro, M. Supporting older adults social network: the design of e-inclusion communication services. Gerontechnology, 7, 153. Fullpaper on CD-ROM edition of the Proceedings of the 6th Conference of the International Society for Gerontechnology, ISG08, 2008

Lewis, T., Langdon, P. and Clarckson, P.J. Cognitive aspects of ageing and product interfaces: Intrface type. In C. Stephanidis (Ed.). Universal Access in HCI, Part I, HCII 2007, LNCS 4554, pp. 731–740, 2007. Berlin Heidelberg: Springer, 2007

LexisNexis. Technology gap survey 2008. http://www.lexisnexis.com/media/pdfs/LexisNexis-Technology-Gap-Survey-4-09.pdf (20.5.2011)

Lievrouw, L. and Livingstone, S. The Handbook of New Media. London: SAGE, 2002 – 2004 reprinted

Lim, C. Designing inclusive ICT products for older users: taking into account the technology generation effect. Journal of Engineering Design, 21 (2), 2010, 189-206. First published November 11 2009. http://dx.doi.org/10.1080/09544820903317001. (18.6.2011)

McMullin, J.A., Duerden Comeau, T. and Jovic, E. Generational affinities and discourses of difference: a case study of highly skilled information technology workers. The British Journal of Sociology, 58(2), 2007, 297-316

Melenhorst, A.-S. Older adults' motivated choice for technological innovation: Evidence for benefit-driven selectivity. Psychology and Aging, 21 (1), 2006, 190

Morris, MG., Venkatesh, V. and Ackerman, PL. Gender and age differences in employee decisions about technology: An extension to the theory of planned behavior. IEEE Transactions on Engineering Management 2005;52(1) (2005), 69-84

Prendergast, D. and Roberts, S. Communication – Practice, systems and technology for seniors. Universal Access in the Information Society, 8, (2008), 59-61

Saariluoma, P. and Leikas, J. Life-Based Design – an approach to design for life. Global Journal of Management and Business Research GJMBR Vol.10 Issue 5: June/July (2010), 17-23

Sackmann, R, and Weymann, A. Die technisierung des alltags: Generationen und technische innovationen. Bremen: BMFT, 1994

Shove, E. and Pantzer, M. Consumers, producers and practices: understanding the invention and reinvention of Nordic walking. Journal of Consumer Culture, 5 (1) (2005), 43–64

Shove, E., Watson, M.and Ingram, J. Designing and consuming: objects, practices and processes. Cultures of consumption programme. ESRC & AHRB, 2005

Silverstone, R. Consuming technologies: media and information in domestic spaces. London: Routledge and Kegan Paul, 1992

Stutzman, F. Thoughts about information, social networks, and privacy. http://fstutzman.com/ 2010/02/25/generations-and-technology/ (20.5.2011)

Tapscott, D. Grown Up Digital. How the Net generation Is Changing Your World. McGraw-Hill, 200

The Official Presence of Croatian Higher Education Institutions on Social Networks

Kruno Golubić
University Computing Centre, University of Zagreb
Josipa Marohnića 5, Zagreb, Croatia
kruno.golubic@srce.hr

Summary

New trends constantly emerging on the Internet bring new ways of communication and dissemination of information. Social networks, which have been present on the Internet for more than 15 years, have lately become a topic of every day conversation, not only among individual users but also institutions. The spread of new technologies affecting even higher education institutions is witnessed by a considerable number of researches, papers and discussions about the use of social networks for educational purposes. The key focus of this paper official presence of Croatian higher education institutions, both publicly and privately funded, on social networks. The presence of RSS feeds on web pages of academic institutions was also analysed. RSS was part of this research due to fact that it can be rather easily converted to content on social networks. This paper does not aim to analyse the use of social networks as educational tools. It rather focuses on the ways in which social networks can be used to upgrade the current means of communication within an academic community and between the general public and the academic community. The total of 170 intuitions was researched for the purpose of this paper. The gathering of data was conducted in two sessions. The first session was during the academic year of 2009/10 and the second session during the academic year of 2010/11. A comparison of these two data sets shows certain trends in the acceptance of social networks among the monitored intuitions. Additionally, a comparison of the adoption levels of social networks between publicly and privately funded institutions is given.

Key words: Croatian higher education, social networks, Facebook, LinkedIn, Twitter, RSS

Introduction

Internet based social networks[1] have been around for quite some time now. According to Wikipedia[2] social networks began emerging on the web in the mid 1990s and have become an everyday topic in the last couple of years.

The spread of new technologies affecting even higher education institutions is witnessed by a considerable number of researches, papers and discussions about the use of social networks for educational purposes. Certain topics of academic research are oriented towards the usage of social networks in one aspect of education[3], while others are focused on the usage habits among students and academic staff.[4] This paper, however, does not deal with that aspect of social networks. It rather focuses on the use of social networks as a means of communication between higher education institutions and their students, academic and non-academic staff and other interested parties.

Behling and Klingner say that "Facebook, LinkedIn and Twitter are ways in which people can connect with one another based on similar interests and networks. Facebook started as a way for college students to connect with one another, but has expanded tremendously.

LinkedIn is a more of a professional social networking site allowing connections to be made and professional job histories and consulting to be shared. Twitter is a micro-blogging service that requires users to share their thoughts in 140 characters or less per tweet... But Twitter's global impact has often been underestimated as well as its academic potential".[5]

They also give a few examples how social networks could be used in academic environment:

- Facebook could be used for discussions among students
- LinkedIn could be used for connecting with Alumni
- Twitter could be used for sending announcements[6]

---

[1] In this paper the term "social network" is used to denote Internet based social networks.

[2] Social networking service. 20 June 2011. http://en.wikipedia.org/wiki/Social_network_service (retrieval date 24 June 2011)

[3] Liccardi, I. The role of social networks in students' learning experiences. ACM SIGCSE Bulletin, 39,4 (2007), pp. 224-237.

[4] Matešić, Marija; Vučković, Kristina; Dovedan, Zdravko. Social Software: Teaching Tool or Not? // 2. International Conference The future of information sciences : INFuture2009 / Stančić, H. et. al. (eds). Zagreb : Department of Information Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, 2009. pp. 443-442. http://infuture.ffzg.hr/2009/ papers/5-07%20Matesic,%20Vuckovic,%20Dovedan,%20Social%20software.pdf (retrieval date 28 August 2011)

[5] Behling, Michelle Pulaski, Klingner, Beth Gordon. The Technological Age of Teaching // Teaching Inclusively in Higher Education / Susan C. Brown, Moira A. Fallon (eds). Charlotte : Information Age Publishing, 2010, p. 162.

[6] Ibid.

Social networks can also be used as a means of communication with future students not only in a phase when high school graduates have to decide which college or university to enrol into but it can be also used as a "medium for students to interact before arriving at university".[7] Use of social networks can be also of interest to academic libraries to reach their users. As Gibbons notices, "With 60%, 75%, or perhaps even more of our students using this social networking services as a communication venue, we just cannot ignore them".[8]

Social networks can also be used for communication among scientists. There are even specialised social networks[9] for scientists such as ResearchGate[10] that was created to "promote knowledge sharing between scientists all over the world and is based on the idea that communication between scientists will accelerate the creation and distribution of new knowledge and assure research quality".[11]

As has already been shown, academic institutions can follow different scenarios for usage of social networks. They can also benefit from large marketing potential that lies in them.[12]

Application of communication strategies created for strictly profit oriented industries to academic institutions might seem illogical and impossible if we think of academic institutions strictly as non-profit organizations. However, with new academic institutions, both profit and non-profit, emerging every year the need for their better self-presentation and communication with the public, that is, all those interested in getting information about universities, is on the increase. A larger number of overlapping activities and services also means that the inter-institutional battle to attract students, staff and funds is likely to become fierce. In this context some communication strategies from profit oriented industries can and should be introduced into the academic world. As Brown states "Ultimately, the choice for organizations is a simple one: they either take part in this

---

[7] Griffiths, Philip; Wall, Anthony. Social Media Use by Enrolment Management // Higher Education Administration with Social Media Including Applications in Student Affairs, Enrollment Management, Alumni Relations, and Career Centers / Laura A. Wankel, Charles Wankel (eds). Bingley : Emerald Group Publishing, 2011, p. 50

[8] Gibbons, Susan. The Academic Library and the Net Gen Student: Making the Connections. Chicago: American Library Association, 2007, p. 89

[9] A comprehensive list of such networks can be found in Apendix 1 of "Social Web: Web 2.0 Technologies to Enhance Knowledge Communities". Eric Pardede (ed). New York : Springer, 2011.

[10] ResearchGate is located at URL http://www.researchgate.net/.

[11] Ayanso, Anteneh; Herath, Tejaswini; Lertwachara, Kaveepan. Social Web: Web 2.0 Technologies to Enhance Knowledge Communities. // Community-Built Databases: Research and Development / Eric Pardede (ed). New York : Springer, 2011., p. 11

[12] Zarrella, Dan. The Social Media Marketing Book. Sebastopol: O'Reilly, 2010

conversation or they don't. What they have to realize though is that if they don't participate in these conversations they simply won't go away. The dialog will go on without them".[13] Academic institutions can utilize social networks as one of their public relations channels. With this type of communication channel a very wide range of users can be targeted. It is Rutledge's opinion that "...the world of social networking is wide and diverse. Although early adopters of social networking skew to the young, the trend has now moved into the mainstream with sites for all ages, backgrounds, and interests".[14]

Publicly funded institutions have a larger base of stakeholders[15] than privately founded ones. In the case of publicly funded institution all tax payers can be considered as fund providers and therefore should have easy access to information about the institutions and their work. Perhaps the easiest and fastest way to inform the public today is over the Internet. It is impossible to know the exact number of Internet users in Croatia or anywhere else in the world for that matter. According to the 2009 GfK report 49% of Croatian population older than 15 uses the Internet. Half of those Internet users use social network websites.[16] The International Telecommunication Union (ITU) data [17] on the number of Internet users in Croatia are given in Table 1.

Table 1 Estimated number of Internet users in Croatia according to ITU

Year	Estimated number of Internet users (%)
2000	6,64
2001	11,56
2002	17,76
2003	22,75
2004	30,91
2005	33,14
2006	37,98
2007	41,44
2008	44,24
2009	50,58

Source:http://www.itu.int/ITU-D/ict/statistics/material/excel/EstimatedInternetUsers00-09.xls (retrieval date 24 Jun 2011)

---

[13] Brown, Rob. Public Relations and the Social Web: How to Use Social Media and Web 2.0 in Communications. London, Philadelphia: Kogan Page, 2009, p. 18.

[14] Rutledge, Patrice-Anne. The Truth About Profiting from Social Networking. Upper Saddle River: FT Press, 2008, p. 2.

[15] In this paper the term "stakeholder" is used to denote students, academic and non-academic staff and fund providers

[16] Gdje smo danas u informatičkoj pismenosti? 17 February 2010. http://www.gfk.hr/public_relations/press/press_articles/005364/index.hr.html (retrieval date 25 June 2011)

[17] Estimated Internet users. http://www.itu.int/ITU-D/ict/statistics/material/excel/Estimated InternetUsers00-09.xls (retrieval date 24 June 2011)

Similarly to the number of Internet users, it is also impossible to certainly determine how many users use a particular social network. The list of the most popular social networks published on Wikipedia has around 200 entries and contains an estimated number of users for some of the networks. The three networks that are listed among the ten biggest social networks on the Wikipedia list[18], namely Facebook, Twitter and LinkedIn will be the focus of this paper.

Among the three mentioned networks Facebook is the only one to offer a possibility for obtaining an estimated number of users through an advertising planning tool. On 11 June 2011, an estimated number of Facebook users from Croatia was 1 405 800[19]. Since there is no means of verification that could guarantee accuracy of user data in the process of account creation it is possible to create multiple user accounts without providing true information. Therefore, the previously mentioned number of Facebook users from Croatia should be taken only as a rough approximation. Despite this fact it is safe to assume that Facebook is the most popular social network in Croatia. Consequently, this paper focuses mainly on the use of Facebook among higher education institutions in Croatia. Twitter and LinkedIn are mentioned but not to such an extent.

Given the numbers mentioned above it seems logical to assume that higher education institutions in Croatia can benefit from their presence on social network web sites. It is up to each institution to decide what kind of content would be published and what kind of communication strategies would be used.

However, it should be stressed that the mere presence on social networks is not enough. If social networks are taken to be a new communication channel for public relations some changes have to be made in existing communication strategies. According to Green "Social media are more about personal relationships and one-to-one dialogue, whereas traditional public relations, and particularly media relations, are generally about broadcasting to a wide audience. Social media make communication much more personal and direct".[20] Therefore, setting up a communication strategy is a prerequisite for a presence on social networks. Access to content that is published on certain social networks can have many forms and can be limited to members of that social network or can be accessible to all Internet users. It often depends on the generator of content to decide whether membership to a particular social network or group of people is necessary or not.

---

[18] List of social networking websites. 15 June 2011. http://en.wikipedia.org/wiki/List_of_ social_networking_websites (retrieval date 24 June 2011)

[19] The stated number of users resulted from the query that was retrieved from "Facebook advertising planning tool" at URL http://www.facebook.com/ads/create/. Query was formulated so that all users from Croatia are counted i.e. no age restriction was set.

[20] Green, Andy. Creativity in Public Relations. London, Philadelphia: Kogan Page, 2010, p, 135.

Research

The goal of this research was to see how many Croatian academic institutions have some kind of official presence on social networks. In addition to social networks, RSS (Really Simple Syndication) usage was also a subject of this research as it is a method of information dissemination which is effective and rather easy to implement.

The term "official presence" is used in this paper for those forms of presence on social networks which can be found or accessed through the official web sites of the researched institutions. This criterion was introduced in order to rule out possible cyber-squatting scenarios. However, the downside of such an approach is that some institutions that established a legitimate presence on social networks but contain no links on their official web sites were omitted from this research.

Each social network has certain characteristics that make them more or less suitable for certain usage scenarios. On Facebook there are three forms of representation: pages, groups and profiles. Each of the representation forms has it strengths and weaknesses. According to Zarrella and Zarrella, "Groups are often faster and easier to set up than Pages. They also offer a more personalized and controlled atmosphere for conversation".[21] The authors furthermore stress that "Groups also offer more control over who is allowed in or out, while Pages are open to everyone. A Group can be open just to a particular school or work network, or to all of Facebook. You can also require permission for joining a Group so that everyone must be approved by an admin (but can see some Group content before requesting to join), or you can make the Group completely secret and visible only to those you invite".[22] Smith and Treadaway say that "fan pages allow the same type of interaction as groups but with many more options for customization and personalization".[23] They also say that "because of the flexibility of using HTML code, fan pages can be customized to look similar to a website".[24] It is clear that Facebook pages and Facebook groups are both a good and legitimate way to represent institutions on Facebook. There is no reason why institution could not use both forms of presence on Facebook. Pages can be used to reach a largest possible number of users, and groups can be used for communication among smaller or larger groups of people (e.g. alumni group or project based group).

---

[21] Zarrella, Alison; Zarrella, Dan. The Facebook Marketing Book. Sebastopol: O'Reilly Media, 2011, p. 75.

[22] Ibid.

[23] Smith, Mari; Treadaway, Chris. Facebook Marketing: An Hour a Day. Indianapolis: Wiley Publishing, 2010, p. 101.

[24] Ibid.

The situation with Facebook profiles is, on the other hand, not so clear and straightforward. As Carlos states in his article "many companies have created an institutional or product presence on Facebook using Facebook profiles… Facebook insists that profiles represent real individuals, not couples, groups, nor companies and their products or services… For its part, Facebook has turned a blind eye, deleting company profiles mostly upon receiving complaints.".[25]

The proof of Facebook's awareness of these issues is the provision of the option that allows conversion of a Facebook profile to a Facebook page. More details about the conversion process can be found at the Facebook help page "Converting your profile into a Page"[26]. It is, therefore, recommended to institutions that are currently using Facebook profiles to switch to Facebook pages.

User accounts and their corresponding timelines on Twitter can be public or private. If someone wants to follow an activity of a user or institution with a private profile, they first need to get authorization for access to a timeline.

According to Rutledge LinkedIn is different from other social networks based on the fact "that its audience is almost exclusively professionals looking to achieve professional goals, such as finding a new job, recruiting employees, generating new business, or locating potential business partners.".[27] This statement shows the potential of LinkedIn which can be explored by institutions, students and employees. LinkedIn gives institutions opportunities to both promote themselves and find new opportunities for their further growth and development. Students that are connected with institutions over LinkedIn can be more easily found by head hunters.

Although it may seem at first that RSS has no relevance in relation to social networks this cannot be further from the truth. The first argument in favour of RSS is that it is used to distribute content across the Internet. The second argument is that RSS can be easily used to automate publishing of content on social networks either through mechanisms that are already part of the social networks or with the help of specialized software or web site services[28]. Additionally, if a social network allows inclusion of HTML content RSS feed can be converted into HTML with the help of web based tools and services[29]. This way they

---

[25] Carlos, Sean. Dear company, it's time to convert your Facebook profile into a page before it's too late. 18.04.2011. http://antezeta.com/news/convert-facebook-profile-to-page (retrieval date 15 May 2011)

[26] Converting your profile into a Page. https://www.facebook.com/help/?page=18918 (retrieval date 24 June2011)

[27] Rutledge, Patrice-Anne. The Truth About Profiting from Social Networking. Upper Saddle River: FT Press, 2008, p. 24.

[28] One of such web based services is Twitterfeed located at URL http://twitterfeed.com/.

[29] Some of those web services and tools are listed in article "RSS To HTML - How To Convert RSS Feeds Into Published Web Pages - A Mini-Guide" at URL http://www.masternewmedia.

would gain another channel for dissemination of information. It is important to stress that if content is only converted from RSS feeds to social networks, without any additional actions such as setting of communication strategy, this would form a one way communication channel. It should be mentioned that a reverse process is also possible; that is, to create a RSS feed from the data published on social networks. Therefore RSS is also a topic of interest in this research.

The conducted research consisted of data gathering from web pages of Croatian academic institutions. A list of institutions researched for the purpose of this paper was created from the data obtained from the web site of the state Agency for Science and Higher Education (ASHE). In June 2010, a set of three documents was downloaded from the ASHE web site[30]:

- list of Universities in Croatia
- list of Polytechnics in Croatia
- list of Colleges in Croatia

The web site addresses of all the institutions listed in the documents were obtained through Google search. Since university constituents, such as faculties, art academies, departments and university centres were not included in the published documents, they were collected from the web sites of the researched universities.

Web addresses for the total of 172 institutions[31] were gathered in June 2010. 125 out of 172 institutions are publicly funded while 31 are privately funded. More detailed information on funding sources is given in table 2.

Table 2 Classification of institutions according to the sources of funding (data from 2010)

	Publicly funded	Privately funded
Universities with their constituents	125	5
Polytechnics	13	2
Colleges	3	24
TOTAL	141	31

As it is shown in Table 2, publicly funded institutions make 82% of all the institutions included in the research.

The total of 172 institutions was examined for the purpose of this paper. The gathering of data was conducted in two sessions. The first session occurred

---

org/content_delivery_and_distibution/RSS-to-HTML/RSS-to-HTML-conversion-tools/2007-03-30.htm.

[30] This set of documents was retrieved from URL http://www.azvo.hr/Default.aspx?sec=219 which seized to exist with redesign of ASHE web site in 2011. Updated data from those documents can now be found at URL http://www.azvo.hr/hr/visoko-obrazovanje/visoka-uilita.

[31] In the context of this paper the term "institution" refers to universities together with their faculties, academies of arts, departments and university centres, and to polytechnics and colleges.

during the academic year of 2009/10 and the second session during the academic year of 2010/11.

In the period between June and September 2010, the research of 172 institutional web sites involved an analysis of the following features of their home pages[32]:

- presence of a link to Facebook
- presence of a link to LinkedIn
- presence of a link to Twitter
- presence of a RSS feed or a link to RSS feed

The web site of each institution was opened in a web browser[33] and its home page was inspected for the presence of the links listed above. No automation methods, such as the usage of scripts for HTML analyses, were used in search for the links. This approach was chosen because it seemed to the author to be congruent with the ways regular users use the links according to their visibility on web pages. There are cases when a link placed within a web page is not visible to users either due to an improper HTML code, some other syntax errors within the web page or due to poor design, e.g. if a link and background colours are to similar, one may not see it.

Additional data gathering during this research session was the reason for a rather long time span of the first session. Due to their irrelevance to this particular topic, these data are not presented in the paper.

The same method was applied again in the second session during May 2011. Web pages from two institutions could not be accessed for the second data gathering session. Those two pages belonged to two publicly funded constituents of the University of Dubrovnik. During the 2010 research session those two pages did not have any elements that were in focus of this research so it is considered that their absence from the second session does not have any strong impact on the results of this research. Consequently, the second session included 170 analysed websites. In the cases when web site addresses of some institutions changed, new pages were visited and analysed and the data compared with those from 2010.

Findings

In the period between the two data gathering sessions, the University of Dubrovnik redesigned its web site. On the new web site, at the footer of every department's web page there is a link to the University of Dubrovnik Facebook profile although the departments have no individual Facebook profile, page or group. The Zagreb School of Management is now part of the VERN Group so

---

[32] The term "home page" refers to the first page of web site that was analyzed

[33] Mozilla Firefox 3.6 was used in the 2010 research session, while Firefox 4 was used in the 2011 session

they also share the same Facebook page. Due to the fact that some constituencies of the universities have a joint profile, there is a difference between the total number of links and the number of profiles on Facebook. Table 3 shows comparison of the results for the presence on social networks.

Table 3 Comparison of the results gathered in 2010 and 2011

	Number of links (2010)	Number of links (2011)
Facebook presence	14	32 (26 individual)
LinkedIn presence	0	2
Twitter presence	2	5

When taking into consideration the total number of found links in contrast to the number of individual links the following results can be discerned.
Twenty-six institutions had presence only on Facebook, 16 of them were publicly funded and 10 of them were privately funded. All institutions that had presence on Twitter or LinkedIn also had presence on Facebook. Only a few institutions had presence on more than one social network:
- only one institution had presence on Facebook, LinkedIn and Twitter at the same time; this institution is publicly funded
- only one institution had presence on Facebook and LinkedIn at same time; this institution is publicly funded
- four institutions had presence on Facebook and Twitter at same time; three of those institutions are publicly funded

As it was stated earlier, the number of publicly funded institutions included into this research decreased from 141 in 2010 to 139 in 2011. The number of privately funded institutions remained the same, 31 of them in both sessions.
When looking at relative numbers, the privately funded institutions have more presence on social networks than the publicly funded institutions. Eleven privately funded institutions out of 31 had some kind of presence on social networks; this makes 35 percent of privately funded institutions. Out of 139 publicly funded institutions 21 had some kind of presence on social networks; this makes 15 percent of such institutions.
From these data it is clear that social networks have a greater acceptance level among privately funded institutions than among publicly funded ones.
By analysing forms of presence, based on data from 2011, Facebook shows three forms of representation:
- Facebook pages - 19 institutions use this form of presence
- Facebook groups - 4 institutions use this form of presence
- Facebook profiles - 9 institutions use this form of presence

In total 21 out of 32 institutions that have official presence on Facebook are publicly funded.

According to the 2011 research five institutions that are present on Twitter use public profiles and, therefore, their timelines are accessible to all Internet users. Four out of five institutions that are present on Twitter are publicly funded.

The same research shows that only two institutions had presence on LinkedIn. Both of them use public groups as their form of presence on LinkedIn and they are both publicly funded.

In the case of RSS feeds two sets of gathered data were analysed, one referring to the existence of RSS feed and the other referring to a number of RSS feeds. In 2010, 35 web pages featured an RSS feed or a link to RSS. In 2011 the number of web pages containing a RSS feed or a link to RSS increased to 45. Some pages had several different forms of RSS feeds[34] or they had more RSS feeds with different content. Data for RSS feeds a presented in Table 4.

Table 4 Number of institutions using RSS feeds

Number of found RSS feeds or forms of RSS feeds	Number of institutions (2010)	Number of institutions (2011)
1	14	16
2	16	20
3	1	2
4	5	5
5	1	1
6	0	1

In total, 38 out of 45 institutions that use RSS are publicly founded.

It is evident that the number of institutions using RSS is significantly larger than the number of institutions that established any form of presence on social networks. The gathered data also reveal only a small number of institutions using both RSS and social networks at same time. It is also interesting to notice that the acceptance level of RSS among publicly and privately funded institutions, 27 and 23 percent respectively, is much more balanced than the one for social networks,

Table 5 Number of institution that use both RSS and social networks

	2010 research	2011 research
Both RSS and Facebook presence	5	10
Both RSS and LinkedIn presence	0	1
Both RSS and Twitter presence	1	3

Table 5 shows that there are institutions that could use some methods of content conversions from RSS to social networks that were mentioned earlier in this paper.

---

[34] These different forms include, but are not limited to, RSS in different versions (0.91, 1.0, 2.0), ATOM, and OPML.

According to the 2011 data set only one institution had presence on all three earlier mentioned social networks as well as RSS feed. Two institutions had presence on Facebook and Twitter and RSS feed at same time. Seven institutions had a combination of Facebook presence and RSS feed. An exclusive use of LinekdIn and RSS feed was not noted during this research.

Conclusion

The results of the research lead to the conclusion that the use of social networks among Croatian higher education institutions is not widespread. Comparison of the data gathered in 2010 and 2011 shows that there is an increase in the number of institutions that have presence on social networks. There is still a lot of potential for the researched institutions of higher education to increase their presence on social networks on the Internet and it is reasonable to expect it. In order to see if the trend of growth will continue new data sets should be gathered according to the same methodology in the following years. If necessary, the emergence and acceptance level of new social networks should also be monitored and included in future data gathering sessions. Among the three monitored social networks Facebook is the most represented one. Given the number of Internet users in Croatia and the number of Facebook users from Croatia this result is expected. It can therefore be expected that Facebook will keep the leading position in the following years despite the fact that social networks are emerging on a daily basis.

Depending on their interests and wishes, institutions using RSS could rather easy establish their presence on social networks. Conversion of RSS content into a form appropriate for social networks is not a complicated task. However, it should be borne in mind that if conversion of content is the only thing performed on the site, with no desire to engage into communication with social network users, an important component of social networks will be left out. It is advisable not to enter the world of social networking sites if there is no clear communication strategy. By engaging into the world of social networks institutions should be more open to possible public criticism. Such criticism could lead to problems if no clear guidelines are given on how to handle such situations. One should also bear in mind that positive sides of opening towards online users outweighs possible negative sides. If a critical group of users demanding presence of certain institutions on social networks does not exist, the launching of such a project should be carefully considered and its cost-benefit analysis certainly carried out beforehand.

Future research should also include a focused survey of institutions that already use social networks or are planning to use them in order to examine their users' opinions. It also remains to be seen what the cause of greater acceptance of social networks is among privately funded institutions compared with publicly funded ones. The data that are publicized on their social networks should also be the subject of a content analysis. Based on that information, presence guide-

lines could be created for the institutions that are about to establish their presence on social networks and thus help them avoid potential pitfalls.

References

Ayanso, Anteneh; Herath, Tejaswini; Lertwachara, Kaveepan. Social Web: Web 2.0 Technologies to Enhance Knowledge Communities. // Community-Built Databases: Research and Development / Eric Pardede (ed). New York : Springer, 2011., pp. 3-20.

Behling, Michelle Pulaski ; Klingner, Beth Gordon. The Technological Age of Teaching. // Teaching Inclusively in Higher Education / Susan C. Brown, Moira A. Fallon (eds). Charlotte : Information Age Publishing, 2010, pp. 155-170.

Brown, Rob. Public Relations and the Social Web: How to Use Social Media and Web 2.0 in Communications. London, Philadelphia: Kogan Page, 2009.

Carlos, Sean. Dear company, it's time to convert your Facebook profile into a page before it's too late. 18.04.2011. http://antezeta.com/news/convert-facebook-profile-to-page (retrieval date 15 May 2011)

Converting your profile into a Page. https://www.facebook.com/help/?page=18918 (retrieval date 24 June2011)

Estimated Internet users. http://www.itu.int/ITU-D/ict/statistics/material/excel/EstimatedInternet Users00-09.xls (retrieval date 24 June 2011)

Gdje smo danas u informatičkoj pismenosti? 17 February 2010. http://www.gfk.hr/public_ relations/press/press_articles/005364/index.hr.html (retrieval date 25 June 2011)

Gibbons, Susan. The Academic Library and the Net Gen Student: Making the Connections. Chicago: American Library Association, 2007.

Green, Andy . Creativity in Public Relations. London, Philadelphia: Kogan Page, 2010.

Griffiths, Philip; Wall, Anthony. Social Media Use by Enrollment Management // Higher Education Administration with Social Media Including Applications in Student Affairs, Enrollment Management, Alumni Relations, and Career Centers / Laura A. Wankel, Charles Wankel (eds). Bingley : Emerald Group Publishing, 2011, pp. 49-67.

Liccardi, I. The role of social networks in students' learning experiences. ACM SIGCSE Bulletin, 39,4 (2007), pp. 224-237.

List of social networking websites. 15 June 2011. http://en.wikipedia.org/wiki/List_of_social_ networking_websites (retrieval date 24 June 2011)

Matešić, Marija; Vučković, Kristina; Dovedan, Zdravko. Social Software: Teaching Tool or Not?. // 2. International Conference The future of information sciences : INFuture2009 / Stančić, H. et. al. (eds). Zagreb : Department of Information Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, 2009. pp. 443-442. http://infuture.ffzg.hr/2009/papers/5-07%20Matesic,%20Vuckovic,%20Dovedan,%20Social%20software.pdf (retrieval date 28 August 2011)

Rutledge, Patrice-Anne. The Truth About Profiting from Social Networking. Upper Saddle River: FT Press, 2008.

Smith, Mari; Treadaway, Chris. Facebook Marketing: An Hour a Day. Indianapolis: Wiley Publishing, 2010.

Social networking service. 20 June 2011. http://en.wikipedia.org/wiki/Social_network_service (retrieval date 24 June 2011)

Zarrella, Alison; Zarrella, Dan. The Facebook Marketing Book. Sebastopol: O'Reilly Media, 2011.

Zarrella, Dan. The Social Media Marketing Book. Sebastopol: O'Reilly Media, 2010.

User Behavior in Tagging in the OPAC: The Example of the Faculty of Humanities and Social Sciences Library in Zagreb

Aleksandra Pikić

Faculty of Humanities and Social Sciences Library
Ivana Lučića 3, HR-10000 Zagreb, Croatia
apikic@ffzg.hr

Dorja Mučnjak

Faculty of Humanities and Social Sciences Library
Ivana Lučića 3, HR-10000 Zagreb, Croatia
dmucnjak@ffzg.hr

Summary

Paradigmatic change has occurred in understanding the relationship of Internet users towards the content available on the Web. Unlike the earlier role of users as passive consumers, the new network, Web 2.0 is based on the user-friendly technologies which are focused on collaborative and interactive information services, social bookmarking services, where users can categorize and store their own web links, images, bibliographic records or PDF files. These technologies allow users to independently organize information in the ways which suit them best. Unlike the traditional organizing, where information specialists or authors describe, organize and classify contents, the users of social bookmarking services create metadata themselves using uncontrolled keywords - tags.

Library of the Faculty of Humanities and Social Sciences in Zagreb also allows its users to tag bibliographic records of library materials. In June 2010 the option of adding and browsing tags in the Koha catalogue was plugged in, and the users were informed about this new library service.

The aim of this study is to explore how the Library users tag bibliographic records, i.e. what keywords are used to organize the information in the catalogue. The research problems were the following: which users are the most common taggers, which type of metadata is marked by tags and how do theme tags correspond to the key words in the catalogue?

Key words: social tagging, subject indexing, OPAC, Faculty of Humanities and Social Sciences Library in Zagreb, folksonomy

Introduction

By the mid 2000s a paradigmatic change has occurred in the understanding of the relationship of Internet users towards the content available on the Web. Unlike the earlier role of users as passive consumers, the new network, Web 2.0 is based on the user-friendly technologies which are focused on collaborative and interactive information services[1]. Such technologies are social bookmarking services like Delicious[2], Flickr[3] and CiteULike[4], where users can categorize and store their own web links, images or PDF files. These technologies allow users to independently organize information important to them in the ways which suit them best. Unlike the traditional knowledge organization, where information specialists or authors describe, organize and classify contents, the users of social bookmarking services create metadata themselves using uncontrolled keywords - tags.

Tag can be defined as "any word that defines a relationship between the online resource and the concept in the user's mind"[5]. The process of adding tags is usually called collaborative or social tagging, and the result of this process is a folksonomy. Folksonomies have evolved from the above-mentioned social bookmarking systems. Munk and Moerk define folksonomies as taxonomies created by users who freely create descriptive metadata by tagging documents[6]. The English Wikipedia defines folksonomy as a classification system that has resulted from the collaborative creation and management of tags in order to capture and categorize content. The term was coined by the information architect Thomas Vander Wal in 2003. This neologism is composed of two terms: the term folk (people) and taxonomy, so the literal translation of the concept would be the taxonomy of the people. Folksonomy is not a classification but a flexible horizontal categorization, which consists of associative, but unrelated concepts that users can add and browse without professional supervision. It represents a departure from the traditional classification; the expert categorization of information through the use of a controlled vocabulary. Everybody can tag at any moment on the Internet regardless of language, time or place where they are. Unlike the subject indexing done by the experts, this tagging is more up to

---

[1] Macan, Bojan. Tehnologije Web 2.0 i njihova primjena u knjižnicama – iskustva Knjižnica Instituta Ruđer Bošković s posebnim osvrtom na njezin blog.// Kemija u industriji. 58, 5(2009); 226.

[2] Delicious. [quoted: 2010-09-15] URL: http://www.delicious.com

[3] Flickr. [quoted: 2010-09-15] URL: http://www.flickr.com

[4] CiteULike. [quoted: 2010-09-15] URL: http://www.citeulike.org

[5] Kakali, Constantia; Papatheodorou, Christos. Exploitation of folksonomies in subject analysis. // Library and information science research, 32(2010); 192.

[6] Munk, T.B. ; Moerk, K. Folksonomies, tagging communities and tagging strategies: an empirical study. // Knowledge organization, 34, 3(2007); 116.

date and captures changes when they happen[7]. In addition to these advantages, the tagging system also has serious flaws. Unlike the traditional classification systems and thesauri, there is no expert control over folksonomies, nor any selection criteria or instructions for tagging. All of this generates a large number of homonyms, terms in the singular as well as plural, "sloppy" tags such as misspellings[8] and the danger of different users attributing different meanings to the same tags[9].

Spiteri (according to Lu, Park and Hu) claims that a folksonomy can create added value to public library catalogues since users can organize their personal information space, create additions to the existing controlled vocabularies and create online communities of interests[10].

Kakali and Papathedorou report that, lately, museums, archives and libraries have also allowed their users to tag documents in their catalogues by themselves[11]. Based on added tags in catalogues, the museums The Steve Collaboration and the Penn Museum have been investigating which items in their work of art collections their users consider important and significant. Furthermore, many libraries have upgraded their public web catalogues with the social tagging systems such as LibraryThing for Libraries[12], VuFind[13], Scriblio[14] etc., and in that manner they are enhancing their catalogues. The authors think that user tags can assist librarians in modernizing the vocabularies of their classification systems and that they can reduce the percentage of unanswered user queries.

The Faculty of Humanities and Social Sciences Library in Zagreb also enables its users to tag bibliographic records of the library collection. In June 2010, the option of adding and reviewing tags in the catalogue Koha was included, and the users were informed about the new service offered by the library (Image 1). In order for users to be able to add tags, they must register with their

---

[7] Steele, Tom. The new cooperative cataloging. // Library Hi Tech, 27, 1(2009); 70.

[8] Thomas, Marliese; Caudle, Dana M.; Schmitz, Cecilia. Trashy tags : problematic tags in LibraryThing. // New Library World, 111, 5/6(2010); 225.

[9] Kakali, Constantia; Papatheodorou, Christos. Exploitation of folksonomies in subject analysis. // Library and information science research, 32(2010); 192.

[10] Lu, Caimei; Park, Jung-ran; Hu, Xiaihua. User tags versus expert-assigned subject terms : a comparison of LibraryThing tags and Library of Congree Subject Headings. // Journal of information science, 36, 6(2010); 767.

[11] Kakali, Constantia; Papatheodorou, Christos. Exploitation of folksonomies in subject analysis. // Library and information science research, 32(2010); 192.

[12] LibraryThing for Libraries. [quoted: 2011-05-15] URL: http://www.librarything.com/forlibraries

[13] VuFind. [quoted: 2011-05-15] URL: http://vufind.org/

[14] Scriblio. [quoted: 2011-05-15] URL: http://about.scriblio.net/

AAI@edu[15] user name and password. Tagging records in the library catalogue is done under the supervision of experts because every added tag must be approved by the librarian.[16]

Unfortunately, the Library itself does not use controlled vocabularies for the subject indexing. Only uncontrolled index terms are used for subject analysis of the library's collections by the librarians/subject experts.
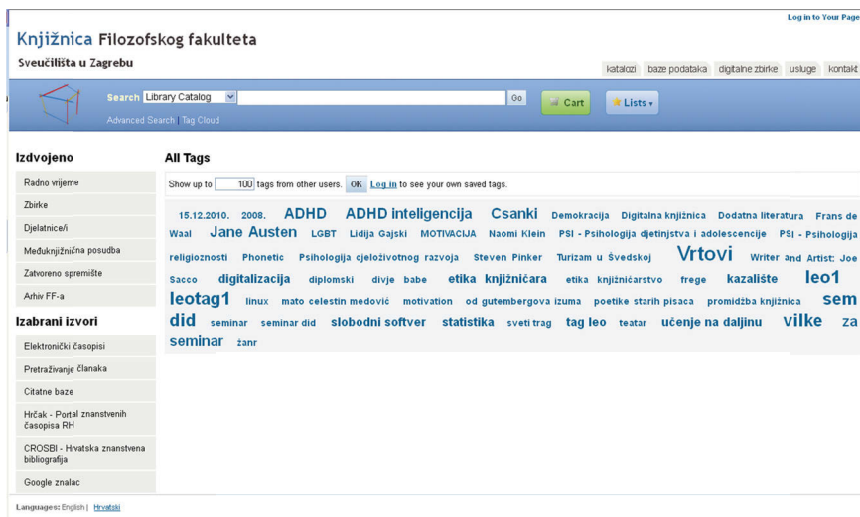

Image 1. Overview of the tag clouds in the catalogue of the Faculty of Humanities and Social Sciences Library

Research on users' behaviour in tagging systems

Numerous studies were aimed at investigating the laws regulating the distribution of tags, the type of metadata represented by the tags as well as semantic links of folksonomies to the knowledge organization system with special emphasis on analyses on users, authors, experts and computer-generated vocabularies. Research has been often conducted on the Delicious service because it is one of the largest folksonomies if we take into account the number of users, but also the number of tags on the Internet[17].

---

[15] AAI @ EduHr is the authentication and authorization infrastructure of the scientific and higher education system in the Republic of Croatia.

[16] For now it is not possible to search the tags in the catalogue. Besides OPAC tagging, the Faculty of Humanities and Social Sciences Library also offers its users creating reading lists and writing comments on and about a particular library item.

17 According to Munk and Moerk, in 2007 more than 100,000 people tagged web sites that they or other people found using their key words in Delicious.

Golder i Huberman (2006) analysed the distribution of the frequencies of tags on Delicious with the aim of discovering regularities in use. The results have shown that the added key words are mostly used for the private purposes of the users, which means that the results have indicated that only that user can make cognitive connections between an individual web site and the added tags.

In late 2005, Munk and Moerk (2007) gathered 76,601 different key words from 500 randomly selected taggers on Delicious. Key words were analysed quantitatively and qualitatively. Quantitative analysis was used to analyse the frequency and the share of individual words by means of statistical correspondence analysis in order to discover possible regularities, while qualitative analysis was conducted on the textual part in order to find different strategies of tagging that the users undertake. The results of the research have shown that:

1. The distribution of tags follow the power law, which means that only a few words are dominant, and the great majority appears only once or a couple of times. The most common are the basic cognitive categories or general categories that all people or people employed in the IT sector have in common.
2. There are three tagging communities or three types of taggers: the well-informed and curious citizen who tags wide categories of culture, the IT expert who tags specific technical categories connected with IT technology and the IT designer who tags terms related to design.
3. Two hundred and forty-five most commonly used tags are situated on the axis from basic social subjects to specific concepts in the IT sector.
4. There are nine broad categories of tagging strategies; from the widest categorizing of content, categorizing of media through format, process categorizing to meta-categorization.

Kakali and Papatheodorou (2010) conducted their research on the data from the catalogue of the library of the Faculty of Social and Political Science at the Pantheion University in Athens, Greece. The researchers wanted to see why and how the students and the faculty tag bibliographic records. They were interested in the number of tags in the singular, plural, in phrases or in one word, acronyms, how many tags there are per record and how much do they correspond to expert subject headings. The results have shown that the tags are dominated by theme metadata, and that the tags for the author, title, editor and geography are a distant second. Furthermore, only 12% of tags have been identical to the subject descriptors of that record. They found that a slightly greater percentage belongs to tags consisting of two or more words, and not one word, and that the majority of tags are terms in the singular. The authors have concluded that the users use tags in order to create a short bibliography on a specific subject.

There has been a series of research conducted on the topic of implementation and tag analysis on the service LibraryThing for Libraries. LibraryThing is a

popular social tagging service used to organize personal book collection[18]. As of June 2011 LibraryThing had over 1,3 million users who added 76 million tags for over 63 million titles of books[19]. Research results have shown that as in Delicious, LibraryThing includes two broadest tag categories: meaningful and personally useful.[20] Lu, Park and Hu (2010) compared tags from the website LibraryThing with experts' subject descriptors in Library of Congress Subject Headings (LCSH). The authors discovered that 97.8% of all the tags cannot be found in the LCSH[21].

This paper is the first to investigate the folksonomy of a library catalogue in Croatia. The aim of this research is to analyse what constitutes the folksonomy of the users of the Faculty for Humanities and Social Sciences Library in Zagreb, i.e. what key words the users use to organize their information in the catalogue. The research problems are as follows: which users are the most common taggers, which type of metadata is marked by tags and how do theme tags correspond to the key words in the catalogue?

Research method

The subjects of this research were the tags which were added by the users to the bibliographic records in the public online catalogue Koha of the Faculty of Humanities and Social Sciences Library. Apart from tags and categories of users who tag, we collected the following data from the bibliographic records: the bibliographic number of the record, the title, the author, co-authors, editors, key words of the subject experts and the type of library materials. The data was collected from June 15th 2010 to June 1st 2011, after which it was exported to a MS Excel table. The table was supplemented with data on the type of tag metadata and with data on the existence or non-existence of information on the subject carried by the tag in the index term field of the bibliographic record. Metadata, as well as information on the previous existence of data, were obtained through content analysis of the tags and the MARC 21 field 653 of the record on which the tag was added. The data was processed and analysed by means of descriptive statistics in the statistical programme package SPSS.

---

[18] Westcott, Jezmynne; Chappell, Alexandra; Lebel, Candace. LibraryThing for libraries at Claremont. // Library Hi Tech, 27, 1(2009); 78.

[19] LibraryThing. Zeitgeist. [quoted: 2011-05-15] URL: http://www.librarything.com/zeitgeist

[20] Mendes, Luiz H.; Quinonez-Skinner, Jennie; Skaggs, Danielle. Subjecting tha catalog to tagging. // Library Hi Tech, 27, 1(2009); 32-33.

[21] Lu, Caimei; Park, Jung-ran; Hu, Xiaihua. User tags versus expert-assigned subject terms : a comparison of LibraryThing tags and Library of Congree Subject Headings. // Journal of information science, 36, 6(2010); 770.

The results and the discussion

In almost a year the users added 147 tags in total to 124 bibliographic records, out of which 48 are different. The great majority of tags were added to one bibliographic record, while only 23 tags were added to 2 records (Table 1). The reason for this relatively small number of tags can be found in the probable perception of the catalogue as still belonging to the librarian-professional domain, which resulted in the users not putting in more effort to engage with the software.

Table 1. Number and percentage of tags per number of bibliographic record

	N	%
1 record	124	84.0
2 records	23	16.0
TOTAL	147	100.0

As it is shown in Chart 1, out of the three types of users, the students are the ones who tag the most. As many as 90% of tags were added by students of the Faculty of Humanities and Social Sciences in Zagreb, while librarians and the faculty have jointly added only every tenth tag. It is obvious that the students, not the academic staff, are the ones who have recognized more the usefulness of this form of knowledge organization.

Chart 1. Percentage of tags created by different types of users (N=147)
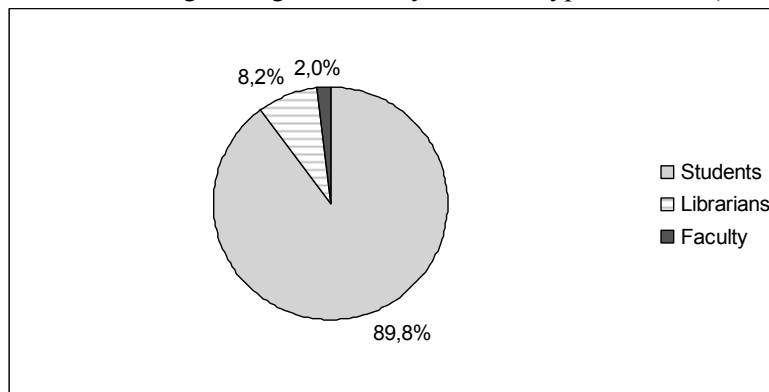
The analysis of the type of documents which were tagged by the users has shown that the books have received the highest number of tags (91.8%), and that the dissertations and theses have received only 8.2% or 12 tags. The other type of documents, such as journals or offprint has not been tagged yet. (Table 2.)

283

Table 2. Number and percentage of tags per type of documents

Type of documents	N	%
Books	135	91.8
Dissertations and Theses	12	8.2
TOTAL	147	100.0

Out of 147 tags, more than half of them are tags with private information (N=77). Private information means that there is no meaningful link between the name of the tag and the information in the bibliographic records. Private information is an example of what was mentioned by Golder and Huberman, and that is that such key words are of service solely to the user for private purposes and have no added value which other users could make use of. A big part of private information from the OPAC of the Faculty of Humanities and Social Sciences Library in Zagreb whose meaning can be surmised, refers to student notes on the materials that they chose for seminar or diploma papers (e.g. tags: za seminar, diplomski, sem did, seminar, seminar did, etc.). Tags carrying private information have been excluded from further analysis.

We wanted to find out which type of metadata is used by the users in order to tag bibliographic records. Content analysis revealed that the tags describe the following metadata: theme, author and the title. The most represented metadata is the theme which appears in three fourths of tags that were analysed (Table 3). Besides the theme, the metadata on the author and the title appear in smaller numbers. The results are similar to those obtained by Kakali and Papatheodourou where the theme dominated among the represented metadata.

Table 3. Number and percentage of tags per different types of metadata

Type of metadata	N	%
Theme	53	75.7
Author	8	11.4
Title	9	12.9
TOTAL	70	100.0

We further analysed only the tags with metadata on the theme since we wanted to find out how they correspond to the key words of subject experts in the bibliographic records. The results showed that in 75.5% cases, the tag of the theme carries completely new information on the subject matter which can be understood as enrichment of the bibliographic record of the tagged material (Table 4). Even though their real number is small (N=44), it can be one of the indicators of the need to revise the indexing system of the Faculty of Humanities and Social Sciences Library.

Table 4. Number and percentage of tags identical to experts' key word on the bibliographic record

Is there an identical experts' key word in the record?	N	%
No	40	75.5
Yes	13	24.5
TOTAL	53	100.0

Conclusion

This is the first research conducted in the library catalogues in Croatia on the subject of folksonomies. The research was aimed to investigate the ways in which users of The Faculty of Humanities and Social Sciences Library organize the knowledge that they need in the catalogue. The selected characteristics of the tags added to the bibliographic library records were also analysed.

The results have shown that the biggest creators of tags are the students of the Faculty of Humanities and Social Sciences and that they mainly use the tags to collect bibliography selected for writing seminar or diploma papers. Out of three types of metadata (author, title, theme), the users mainly write the theme as the tag. An interesting finding was that three fourths of theme tags are not noted in the MARC 21 field 653 of the tagged bibliographic record, meaning that they do not exist in the key words of the subject experts. Apart from this result, the need for revising the indexing system of the library is also indicated by the fact that the Faculty of Humanities and Social Sciences Library does not have a controlled vocabulary for the subject indexing of the bibliographic records.

Of course, we have to keep in mind that the number of added tags is very small and that only a significantly greater number would make it safer to make conclusions about the main tendencies in the folksonomy of the users of this library.

The tags should be taken as the user suggestions of the new terminology or an alternative to already existing concepts.

References

Golder, Scott A.; Huberman, Bernardo. Usage patterns of collaborative tagging systems. // Journal of Information Science. 32, 2(2006); 198–208.

Gordon-Murnane, Laura. Social bookmarking, folksonomies and web 2.0 tools. // Searcher. 14, 6(2006): 26-38.

Kakali, Constantia; Papatheodorou, Christos. Exploitation of folksonomies in subject analysis. // Library and information science research. 32(2010); 192-202.

Lu, Caimei; Park, Jung-ran; Hu, Xiaihua. User tags versus expert-assigned subject terms : a comparison of LibraryThing tags and Library of Congree Subject Headings. // Journal of information science, 36, 6(2010); 763-779.

Macan, Bojan. Tehnologije Web 2.0 i njihova primjena u knjižnicama – iskustva Knjižnica Instituta „Ruđer Bošković" s posebnim osvrtom na njezin blog. // Kemija u industriji. 58, 5(2009); 226-228.

Mendes, Luiz H.; Quinonez-Skinner, Jennie; Skaggs, Danielle. Subjecting tha catalog to tagging. // Library Hi Tech, 27, 1(2009); 30-41.

Munk, Timme Bisgaard; Moerk, Kristian. Folksonomies, tagging communities and tagging strategies: an empirical study. // Knowledge organization. 34, 3(2007); 115-127.

Peterson, Elaine. Beneath the Metadata : Some Philosophical Problems with Folksonomy // D-Lib Magazine. 12, 11(2006). [quoted: 2010-09-15] URL: http://www.dlib.org/dlib/november06/peterson/11peterson.html

Steele, Tom. The new cooperative cataloging. // Library Hi Tech, 27, 1(2009); 68-77.

Thomas, Marliese; Caudle, Dana M.; Schmitz, Cecilia. Trashy tags : problematic tags in LibraryThing. // New Library World, 111, 5/6(2010); 223-235.

Westcott, Jezmynne; Chappell, Alexandra; Lebel, Candace. LibraryThing for libraries at Claremont. // Library Hi Tech, 27, 1(2009); 78-81.

Twitter as a Structured Information System

Dorja Mučnjak

Faculty of Humanities and Social Sciences Library
Ivana Lučića 3, HR-10000 Zagreb, Croatia
dmucnjak@ffzg.hr

Aleksandra Pikić

Faculty of Humanities and Social Sciences Library
Ivana Lučića 3, HR-10000 Zagreb, Croatia
apikic@ffzg.hr

Summary

In the past few years Twitter has achieved a tremendous success. Twitter is an online microblogging service, the tweets up to 140 characters allow people to use Twitter via SMS platforms or computers. With more than 200 million users and 155 million posted messages a day, Twitter has become a vast storehouse of content, i.e. information and knowledge. Such large amount of data that can be accessed on Twitter requires a structure in order to help users find information they need.
The purpose of this paper is to explore ways of structuring information on social networks, especially on Twitter. The aim of this study is to search for and present the most popular methods of gathering and structuring data published on the social networking site Twitter.

Key words: information and knowledge management, Twitter, hashtags, Twitter aggregators, Twitter directories

Introduction

In the past few years Twitter has achieved a tremendous success. As one of the three most popular social networking sites (Facebook, Twitter, LinkedIn), it has evidently found the path to its users. Twitter is an online microblogging service, the tweets up to 140 characters allow people to use Twitter via SMS platforms on most mobile phones, but the tweets can also be posted from computers – one of the reasons Twitter is so popular and widely used. Since the 140-character tweets do not allow extensive posts (especially in comparison to blogs where posts can be much longer), the users quickly realized that Twitter could be used as an ideal pointing device to other websites, for sharing links to longer articles, blog posts, audio-video materials, etc. Websites that were once dominantly accessed from Google search results are now seeing a growing number of new

visitors coming from shared links on Facebook or Twitter. In short, Twitter is a social networking site, a live searching tool and a link sharing place.

With more than 200 million users and 155 million posted tweets per day, Twitter has become a vast storehouse of content, i.e. information and knowledge. Such large amount of data that can be accessed on Twitter requires a structure in order to help users find information they need.

The purpose of this paper is to explore ways of structuring information on social networks, especially on Twitter. The aim of this study is to research and present the most popular methods of gathering and structuring data published on the social networking site Twitter.

Information and knowledge management

Today's society is a knowledge society. New data are being generated in a vast amount on a daily basis.

In order to transform this large amount of data into knowledge, first we have to contextualize the data, i.e. transform them into information and then, through the process of description and evaluation, transform the information into knowledge.[1]

The knowledge management is the most important means of prosperity. In the knowledge society, information and knowledge management, as a concept, cannot be limited only to business intelligence gathering, but these processes should be integrated into the (re)production of everyday knowledge.[2]

In the last few decades, the information and communication technology (ICT) has been a great help in the knowledge management, i.e. in dissemination and (re)generation of knowledge. However, it must be said that ICT is a very useful tool, yet not the resource itself. The dissemination and generation of knowledge is primarily an interaction between people, and ICT is a means of transfer of knowledge.[3] But, integration of ICT in everyday life has evidently changed the ways of communication, generation and dissemination of knowledge. For example, many people see Google as a unique source of information, it has become most visited "public library" without librarians as intermediaries who have skills and knowledge of retrieving, gathering, evaluating, and disseminating information.

At this point the educational process is no longer a process where a teacher transfers his knowledge to students. Today's education[4] is a process of

---

[1] Afrić, Vjekoslav; Lasić-Lazić, Jadranka; Banek Zorica, Mihaela. Znanje, učenje i upravljanje znanjem. // Odabrana poglavlja iz organizacije znanja / Lasić-Lazić, J. (ed.). Zagreb : Zavod za informacijske studije, 2004; 37.

[2] Ibidem; 57.

[3] Ibidem; 56.

[4] Here, the term education is used for formal, non-formal and informal learning.

knowledge (co)production, i.e. a teacher and his students together produce the knowledge through a research process.[5] Apart from teachers, a community also plays a significant role in the production of knowledge. The knowledge production and organization within community of peers has no longer a top-down, but a bottom-up structure.[6] Such a structure is the result of collaboration and communication. A social network of peers is a productive generator of information, but a key question is how to organize information to make it usable and actual knowledge.[7]

Social networking sites – communication and collaboration tools or sources of information?

Everyday technological changes and progress modify the production and transfer of knowledge. In the past few years online social networking sites (*SNS*) have facilitated communication within social networks. Social networking sites are low-cost easy-to-use tools that allow people to share and use a lot of information quickly and easily.

As communication and collaboration tools, these sites are structured as personal (or egocentric)[8] with an individual in the center of their own community; they are organized primarily around people not around interests. That means that one is a "friend" or a "follower" of a person one share interests with – the foundation of this relationship is trust, this is the point of difference at forums that are organized around interests and where trust is not that important because participants are strangers. In that sense, social networking sites are great communication and collaboration tools that allow people to interact with the element of stronger confidence in data.

As said before, social networks in the real world play a significant role in the knowledge (co)production – participants share information with friends, colleagues, relatives… In virtual online world the situation is the same, but facilitated with tools (as social networking sites) that allow people to create larger communities. Yes, social networks in the real world create a source of information, but a mass interaction between people on social networking sites that

---

[5] Afrić, Vjekoslav. Upravljanje sadržajem učenja i znanja. // Odabrana poglavlja iz organizacije znanja / Lasić-Lazić, J. (ed.). Zagreb : Zavod za informacijske studije, 2004; 63-64.

[6] Avram, Gabriela. At the Crossroads of Knowledge Management and Social Software. // The Electronic Journal of Knowledge Management. 4(2006), 1; 1.
http://www.ejkm.com/issue/download.html?idIssue=8 (20.5.2011.)

[7] Upravljanje sadržajem učenja i znanja. 2004; 64.

[8] Boyd, Danah M.; Ellison, Nicole B. Social network sites : definition, history, and scholarship. // Journal of Computer-Mediated Communication. 13(2008), 1; 219.
http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html (27.5.2011.)

produce a vast amount of data makes an even more considerable source of information.

To cope with this great amount of data, one needs guidelines. Depres i Chauvel[9] proposed a framework for categorizing five types of knowledge management activities that could be easily used when retrieving information on social networking sites:

- Scan/map - pointing to the world of overview of data;
- Acquire/capture/create – associated with the world of research, development and creation;
- Package/codification/representation/storing – related to the world of databases, information and knowledge bases, organizational memory;
- Apply/share/transfer – related to the world of competencies, teamwork, intranets and cross border sharing;
- Reuse/innovate/evolve/transform – associated to the world of leverage, intellectual assets and innovation.

Aforementioned activities are mainly meant for organization of business intelligence, but these principles could be used for organization of information of any kind and in any environment. Evidently, social networking sites could be used in any of these five processes:

Scan/map – social networking sites are an ideal place to provide information on interesting people, data, projects, etc. For example, following people or organizations gives us an overview of information they share with their friends/followers.

Acquire/capture/create – browsing social networking sites, e.g. Twitter, is mainly focused on discovering the needs, expertise and offers of other people and indirectly of other organizations.

Package/codification/representation/storing – social networking sites, as a storage of content, enable to find, collect and store information immediately and, not less important, without intermediaries.

Apply/share/transfer – on social networking sites, one can easily share information.

Reuse/innovate/evolve/transform – information posted on social networking sites remain available for a longer period of time (if not forever) and they could be meant for everybody (in case of public access) or only for members of a particular group (in case of limited access).[10]

In the beginning social networking sites were primarily communication and collaboration tools. However, different features that social networking sites offer to users (publishing videos, photos, other multimedia, etc.) have transformed them into a significant source of content.

---

[9] At the Crossroads of Knowledge Management and Social Software. 2006; 5.

[10] Ibidem; 5-7.

But, it has also become evident that the great amount of information generated on social networking sites needs some kind of structuralization in order to facilitate retrieving information.

Three Twitter's characteristics

In this paper, we speak mainly about Twitter – the online microblogging service with more than 200 million registered users who post more than 155 million *tweets* (140-character posts) a day.[11]

After a big success of Facebook, the concept of Twitter as a microblogging service could have sounded weird. Facebook, apart from being a social networking tool, offers interesting multimedia features, such as posting photographs, creating photo-albums, posting videos, playing games or using different applications designed for fun. In addition, on Facebook one can update his/her status, i.e. one posts a note about himself/herself, about what he/she is doing, what he/she is thinking about… Unlike Facebook, Twitter allows posting only short notes, called *tweets*.

Twitter was originally created as a communication platform that allows posting short massages (*tweets*) as an answer to the question *"What are you doing?"*, just like updating the status on Facebook (but limited to 140 characters). Without the possibility to post other interesting objects (photos, videos, music…), Twitter users had soon discovered its simplicity and straight-forward nature and came up with an idea how to make it interesting. Suddenly, it became a great tool to say what was happening around you, to comment on daily news, to inform others about important discoveries... Since the length of tweets is very limited, people started to use Twitter dominantly as a **link-sharing tool**, as a pointing device to other objects – websites, discussions, blogs, etc. To shorten long URLs, people started to use online URL shortening services (for example: http://bit.ly, http://tinyurl.com). Short URLs allow otherwise long web addresses to be referred to in a Tweet. The company itself recognized the way users had used Twitter, so, a few years after its launch, they decided to replace the familiar *"What are you doing?"* tag line that had sat atop status update box with a tag line *"What's happening?"*

Created primarily as a **social networking site**, Twitter allows users to *follow* (to receive other people's tweets on the profile) other users (people, institutions, firms…) as *followers*. Users can choose whether they want their accounts to be public or protected. In case of a public account, profile pages would be visible to everyone. A protected account allows only approved followers to view that

---

[11] What is Twitter? http://business.twitter.com/basics/what-is-twitter (25.5.2011.)

account's tweets. The average Twitter user follows a few dozens of people: a collage of friends, colleagues, and a handful of celebrities.[12]

Twitter has always been very easy to use not only on computers but also via SMS platform. Twitter itself is a very simple program. However, it allows third party programmers to develop different useful applications – for computer and for mobile device use (especially BlackBerry and iPhone). Such applications make Twitter more user friendly, contributing to its popularity.

Evidently, Twitter is a well-used communication tool, or social networking device. But the number of 155 million of Tweets a day shows that Twitter is also a great source of content. It has become important information storage and knowledge generator.

Twitter has another interesting feature: a search box that gives a **real-time** view onto the conversation of just about any topic imaginable. On the profile page, Tweets are being updated in a real time. This makes Twitter an ideal search engine for the newest events. It is also the reason why a certain number of websites, once dominantly accessed from Google search results, now see a growing number of new visitors coming from shared links on Twitter.[13]

To use Twitter as a source of information more easily, it is necessary to structuralize its content. There are a number of community conventions (without any intervention in program) for grouping topics, events or people by certain symbols.

One of the most used community conventions is a *hashtag (#)*, the symbol used to mark relevant keywords or topics in a Tweet, i.e. to categorize and join all tweets with the same topic and show them easily in Twitter Search. If one uses a hashtag on a public account, anyone who does a search for that hashtag may find it.

Another common convention for organizing Tweets is the symbol @ (*at*). It can be used to directly address other user (convention *@Reply*) – @username in the beginning of a Tweet; or to mention other user (convention *@Mention*) – @username anywhere in a Tweet.

These conventions are not always sufficient for structuralizing content. Fortunately, there is a great number of third party applications that can help retrieve information more successfully. The most popular ways of structuralizing content published on Twitter are Twitter aggregators and directories.

---

[12] Johnson, Steven. How Twitter Will Change The Way We Live (in 140 characters or less). // Time. 173(15.6.2009), 23; 34-35. http://www.time.com/time/business/article/0,8599,1902604, 00.html (20.5.2011.)

[13] Ibidem.

Social network aggregation

Social network aggregation is a process of collecting content from multiple social networking services, such as Twitter. The aggregation tasks are most frequently performed by social network aggregators, the tools that pull together information into a single location or help a user to consolidate multiple social networking profiles into a single one. Various aggregation services provide tools or widgets which allow users to collect massages, track friends, gather bookmarks, search across multiple social networking sites, read rss feeds for various sites, access their profiles from a single location, etc.[14]

Another way of collecting tweets of various users is through link directories. Web directories or link directories list links to other websites organized by category and subcategory. Most of the directories are general, but some niche directories focus on restricted regions, single languages, or specialist sectors. They can also be used for collecting profiles of various Twitter users or trending topics on Twitter.

The topic of this paper are aggregators that collect tweets of various Twitter users in a single location and link directories of Twitter profiles or topics. The focus of our interest will not be on aggregators that consolidate multiple social networking profiles into a single one, but on the aggregators that pull together Tweets into a single location.

In the previous sections we have presented the necessity to structure the content on Twitter.

Apart from the Twitter community conventions of # (hashtags) and @ (Mentions) used in tweets, the aggregation of content is another way of structuralizing information on Twitter. Twitter allows third party programmers to create different applications for Twitter, and many aggregation tools are made by third parties. So, it must be pointed out that the content published on Twitter is mainly not structuralized by Twitter tools but by third party applications or community conventions.

Twitter aggregators and directories

Since Twitter is a rather new social networking tool, there is only a small number of scholarly articles or other literature on the subject. However, there are some guidelines that can help find information on Twitter content aggregation such as blogs,[15, 16] directories,[17] and web portals.[18]

---

[14] Social Network Aggregation. Interpiseo. URL:
http://www.interpriseo.com/resources/general_info_articles/Social%20Network%20Aggregation.pdf (20.5.2011.)

[15] Odden, L. 12 Twitter Stream Aggregators To Make You Smarter. / TopRank : Online Marketing Blog. URL: http://www.toprankblog.com/2009/02/12-twitter-stream-aggregators/ (20.5.2011.)

In terms of aggregation of content published on Twitter, there are several ways of collecting information: directories, aggregators, popularity counters, and newspapers. Some of these directories or aggregators are capable of processing both functions at the same time: listing and aggregating.

Twitter directories are lists of links to Twitter accounts organized by certain criteria: by people (very common among politicians) or by subject (list of Twitter profiles posting tweets about the same subject). There are also directories that collect links to other Twitter directories or aggregators.
Examples:
blog.govtwit – list of Twitter aggregators on politics
Just Tweet It – organized by subject
U.S. Air Force – list of Twitter profiles on U.S. Air Force

Twitter directories/aggregators – as mentioned earlier, some of these applications have two functions – collecting links to individual Twitter profiles and aggregating tweets posted on those profiles. Many politicians have discovered these tools and use them frequently.
Examples:
GovLuv – U.S. government representatives
politter.com – Japanese politicians
poliTwitter.ca – Canadian politicians
politwitter.de – German politicians
TweetCommons – Canadian politicians
TweetCongress – U.S. congressmen

Twitter aggregators automatically harvest tweets on different subjects from different Twitter accounts.
Examples:
Europatweets – aggregator of tweets on EU
Retailer Twitter Aggragation – aggregator of tweets on retaile
Retweetist – aggregator of retweets

---

[16] Worth, Jon. Who are the main players in the EU-twittersphere? / Jon Worth Blog. 27.01.2010. http://www.jonworth.eu/who-are-the-main-players-in-the-eu-twittersphere/ (25.5.2011.)

[17] Lunceford, Steve. 50+ government Twitter and social media directories. / Blog.govtwit. 25.10.2010. http://www.blog.govtwit.com/2010/08/25/50-government-twitter-and-social-media-directories/ (20.5.2011.)

[18] Finn, Greg. 3 Aggregation Tools For Tracking Tweets. 21.9.2009. http://searchengineland.com/tracking-tweets-how-to-twitter-aggregation-tools-22699 (20.5.3011.)

Tweetmeme – aggregator of tweets by subject
Twistori – aggregator of tweets containing one of the following words: love, hate, think, believe, feel, wish
TwitLinks – aggregator of tweets on technologies

Popularity counters
Twitter **popularity counters** aggregate the most popular tweets.
Examples:
TwitterCounter
Twiturls
Twiturly
Twopular

Newspapers
The most interesting case of aggregation is the site that organizes links shared on Twitter into an easy newspapers-style format, and can be created for any Twitter user, list or hashtag (#).
Examples:
paper.li

Conclusion
Twitter is a well-used link-sharing and live-searching social networking site. With more than 155 million of tweets a day it has become a great source of information. Twitter is a very simple tool which does not offer possibility of structuralizing this vast amount of data, but it allows users and third parties to create different applications that can help structuralize the content.
The most popular ways of structuralizing the content on Twitter are the community conventions: the hashtags (#) and the @ symbol, which pull together all the tweets with these symbols. The hashtags mark specific topics and the @ symbol helps to gather all the tweets a person was mentioned in.
Twitter has a very simple and straight-forward nature; it does not offer many different ways of structuralizing information but it allows others to create interesting applications. Many of its users took this opportunity and have created many different applications that harvest information from Twitter and structuralize the content. Most common applications are directories (lists of links organized by some criteria), aggregators (tools that pull together information into a single location), popularity counters (organized by criteria of popularity of tweets), and newspaper-style sites (sites which organize links shared on Twitter).
All these useful applications and conventions have grown out of the necessity of structuralizing the great amount of information published on Twitter.

References

Afrić, Vjekoslav. Upravljanje sadržajem učenja i znanja. // Odabrana poglavlja iz organizacije znanja / Lasić-Lazić, J. (ed.). Zagreb : Zavod za informacijske studije, 2004; 62-101.

Afrić, Vjekoslav; Lasić-Lazić, Jadranka; Banek Zorica, Mihaela. Znanje, učenje i upravljanje znanjem. // Odabrana poglavlja iz organizacije znanja / Lasić-Lazić, J. (ed.). Zagreb : Zavod za informacijske studije, 2004; 33-61.

Avram, Gabriela. At the Crossroads of Knowledge Management and Social Software. // The Electronic Journal of Knowledge Management. 4(2006), 1; 1-10. http://www.ejkm.com/issue/download.html?idIssue=8 (20.5.2011.)

Boyd, Danah M.; Ellison, Nicole B. Social network sites : definition, history, and scholarship. // Journal of Computer-Mediated Communication. 13(2008), 1; 210-230. http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html (27.5.2011.)

Finn, Greg. 3 Aggregation Tools For Tracking Tweets. 21.9.2009. http://searchengineland.com/tracking-tweets-how-to-twitter-aggregation-tools-22699 (20.5.3011.)

Johnson, Steven. How Twitter Will Change The Way We Live (in 140 characters or less). // Time. 173(15.6.2009), 23; 32-37. http://www.time.com/time/business/article/0,8599,1902604,00.html (20.5.2011.)

Lunceford, Steve. 50+ government Twitter and social media directories. / Blog.govtwit. 25.10.2010. http://www.blog.govtwit.com/2010/08/25/50-government-twitter-and-social-media-directories/ (20.5.2011.)

Odden, Lee. 12 Twitter Stream Aggregators To Make You Smarter. TopRank : Online Marketing Blog. http://www.toprankblog.com/2009/02/12-twitter-stream-aggregators/ (20.5.2011.)

Social Network Aggregation. Interpiseo. http://www.interpriseo.com/resources/general_info_articles/Social%20Network%20Aggregation.pdf (20.5.2011.)

What is Twitter? http://business.twitter.com/basics/what-is-twitter (25.5.2011.)

Worth, Jon. Who are the main players in the EU-twittersphere? / Jon Worth Blog. 27.01.2010. http://www.jonworth.eu/who-are-the-main-players-in-the-eu-twittersphere/ (25.5.2011.)

Hyperlinks

blog.govtwit http://www.blog.govtwit.com (30.5.2011.)
Europatweets. http://www.europatweets.eu/ (30.5.2011)
GovLuv http://govluv.org/ (29.5.2011.)
Just tweet it http://justtweetit.com/ (30.5.2011.)
Paper.li http://paper.li/ (30.5.2011.)
Politter.com http://politter.com/en/ (30.5.2011.)
poliTwitter.ca http://politwitter.ca/ (30.5.2011.)
Politwitter.de http://politwitter.de/ (30.5.2011.)
Retailer Twitter Agreggation http://www.noturnonred.org/twitter/ (30.5.2011.)
Retweetist http://retweetist.com/ (30.5.2011.)
Social Media. U.S. Air Force. http://www.af.mil/socialmedia.asp (30.5.2011.)
Tweermeme http://tweetmeme.com/ (30.5.2011.)
TweetCommons http://ca.tweetcommons.com/polis/running (30.5.2011.)
TweetCongress http://tweetcongress.org/channels/view/us_media (30.5.2011.)
Twistori http://twistori.com/ (30.5.2011.)
TwitLinks http://twitlinks.com/ (30.5.2011.)
Twitt(url)y http://twitturly.com/ (30.5.2011.)
Twitturls http://twitturls.com/ (30.5.2011.)
Twopular http://twopular.com/ (30.5.2011.)
TwitterCounter http://twittercounter.com/ (28.5.2011.)

MEDIA COMMUNICATION

From Who and What to How and Why – The Future of Online Encyclopaedias

Nenad Prelog
University of Zadar
dr. Franje Tuđmana 24i, Zadar, Croatia
nenad@edemokracija.hr

Domagoj Bebić
Faculty of Political Science, University of Zagreb
Lepušićeva 6, Zagreb, Croatia
domagoj@edemokracija.hr

Summary

This paper discusses changes in the structure of knowledge, the increasing demand for abilities of search and retrieval, assessment and evaluation, organization and creative use of relevant information. It is a review of the topic intended to serve as a basis for further research aimed at answering the many questions that arise from this paper. Today, users are finding content through search engines. This requires a different approach to the organization of encyclopaedias and other lexicographical issues. All can be found, but it is also important to know where and how to look for it.

The conducted research centered around the quality of "coverage" of some, in Croatia well known, lexical units, in different Wikipedias: four regional languages and editions (Croatian, Bosnian, Serbian, Serbo-Croatian), and four world languages (English, French, German and Spanish). We observed the presence of writers and athletes in those languages. None of the selected writers had an article in every language/edition. For instance our writers are almost non-existent in Spanish and French editions, and the situation is only slightly better when it comes to German. The representation of athletes was much better, almost all of them occurring in the selected world languages.

At the end we come to a few questions, such as whether it is more important to write "for yourself", i.e. to work on creating the best possible encyclopaedia intended for audiences in Croatia, and all those who use the Croatian language, or should we systematically work on the presentation of "our" issues in the publications in other languages; whether it be persons or texts from the sphere of politics, history, or (most widely understood) culture and art. If we agree that we will get relatively successful answers to questions starting with who, what, where, when, and much less successful to those beginning with how or why, then this answers the question on the role of the online encyclopaedia in

the transfer of knowledge. The value of the knowledge contained in the answer to the question posed with how or why is "value added" to the encyclopaedia; it is what distinguishes it from a dictionary or a search engine.

Key words: online, encyclopedia, Wikipedia, web, search engine

Introduction: Information is a search

The digitalization of all aspects of life and work has had a huge impact in the media, because it changed the former (more or less one-way) model of production and transfer of information from the source (author, publisher) to the user (audience), at the same time overcoming the increasing number of (emerging) information, by using new methods and techniques of search and retrieval, where often the simplest act of connecting concepts (sending, linking) is becoming more important than the information content itself. [1]

This is particularly evident in the organization and the presentation of knowledge. Because of the constant changes in its structure, the relationships between facts, definitions, understanding, context, the history and future of an event or phenomenon, of a person and its environment, etc. are changing. At the same time the importance of learning the facts is declining, and increasingly is being replaced with the ability to search and find, assess and evaluate, organize and creatively use relevant information. Since the form of learning is changing – it is less about learning the facts and more and more about learning the ways to use them - organized facts, available to everyone are essential for the functioning of the educational system.

Likewise we need to differentiate the public knowledge accessible to all and free of charge from the so-called private knowledge, knowledge that is someone's property, knowledge that is somehow protected, by copyright or other rights (e.g. commercial law). Nowadays, the content itself is often not protected, but the way in which the content is organized is protected [2] (e.g. data from the phone book are not protected, but the directory as a database is protected) or the reproduction of a public content is protected etc.

In a society that is increasingly based on the use of universally available, mostly free sources of information on the Internet, it becomes necessary to fundamentally redefine the role of those who "produce knowledge". Today the emphasis is on searching and finding information, the emphasis is not only on the content but even more on links, or references between the content, expanding the origi-

---

[1] Prelog, N: *Elektroničko (interaktivno) nakladništvo danas – Kako je mreža poosobila masovne medije (II),* Medijska istraživanja, no. 1, 1999. http://www.mediaresearch.cro.net/clanak.aspx?l=hr&id=126

[2] Smedinghoff, T: *Legal Guide to Multimedia.* Adison-Wesley Publishing, Reading, 1994.

nal context in which the requested concept appears. Finally, what is asked for is a dynamic rather than a static approach, constant updating and alignment with the actual situation. If the documents (texts or pictures) are not updated for a longer time the frequency of their use reduces, and the number of users and the influence of the site in the process of learning and exchanging knowledge rapidly decrease.

Electronic editions of the most famous encyclopaedias appeared at the end of the twentieth century. Among the first was the Academic American Encyclopedia, as early as 1985 and in the early nineties Compton's Encyclopedia - in 1992, first on CD, and later on DVD. Online editions followed a few years later, leading with Britannica Online in 1994. In the beginning the role of the Internet came down to only the possibility to update content/releases on a permanent, yet computer-readable medium. It was only later followed by "real" online encyclopaedias, first as copies of electronic editions with the addition of interactivity. In the early 21st century complex knowledge databases begin to dominate, which - building on the multimedia capabilities and the comprehensiveness of the web - provide more information and a large number of services to its users in this respect surpassing in many ways the conventional editions, both on paper as well as those on DVD.

Today, most online editions of encyclopaedias include a number of additional content and possibilities of getting the requested information, including dictionaries, thesauruses, atlases, lists of important quotations, different Happened on this day categories, historical overviews of events or the developments of some ideas, lists of acronyms, abbreviations and idioms, current news, statistics and a range of other content such as important historical documents, famous speeches, excerpts from literary works, automatic translators, quizzes, blogs and so on and so forth.

With the increased importance of new information technologies and the use of new media to store and transfer information, the structure of information products (this includes all artistic, scientific and other works), and the way of presentation of such content changed. It's not just about the conversion of data from one medium to another; an electronic encyclopaedia assumes a different organization of information and interactivity in work (the set of information found is adjusted to the current user request). Encyclopaedias online strive to the total access to knowledge, allowing the exchange of ideas, they also protect cultural heritage and particularities, and constantly explore new possibilities for connecting (both existing and new) content.

The limitations of search engines

With the appearance and the growing importance of information search engines (primarily Google) the mode of access to information contained on the Web quickly began to change. While a few years ago it was common when reading online editions of newspapers or magazines to start from the first, front page,

and then - though rarely - to advance page by page, imitating the conventional edition or - much more frequently – to look up in the index, or the review of sections those titles which we are interested in and then select the article for reading or viewing or listening, it is no longer the dominant approach. Part of the users will reach the information they need (text, image or video) with the help of a mediator, a portal devoted to news services, or dissemination of information to defined areas of interest (RSS). However, today the vast majority of users come to that which they are interested in with the help of search engines, not taking into account the context in which this information is published, often not even noticing the title of the newspaper or the magazine that brings them. Thus it is clear that encyclopaedias on the Web cannot behave as if they are self-sufficient, assuming that the future users will start searching from the home page of the publication or the publisher. The tools offered to users in order to facilitate finding the required information are necessary, but their importance comes into play only when the user begins his journey to knowledge from the home page - which happens less often - or when he is already on the pages of the encyclopaedia/lexicon/vocabulary, etc., regardless of the way he got there.

On the Web, almost everything is possible to find but it is completely different if you have to know where and how to look, while today we are accustomed to simply type a few words into some of the browsers. This naturally results in certain mental laziness, but few people are concerned about this because the goal, in this case the answer to these questions, often justifies the means. It has been shown many times, that even Google cannot answer all his questions. If we take as a criterion of quality those 5 eternal journalistic questions (who, what, where, when, why) we easily find that the answers to the questions asked with the first four interrogative pronouns are always relatively accessible, even adequate, accurate and sufficient, while things mostly get stuck when you ask a question with **why**.

Demands set on the encyclopaedia may have more ambiguity, unexpectedness and lack of adaptation to the system of finding information, than its developers would expect. The encyclopaedia is not a directory in which the amount of questionability is predictable, where the name, address, the name of company/organization and/or branch of industry unambiguously determines the required number. The amount of ambiguity in any remotely ambitious collection of knowledge is far greater and therefore the description of individual units (tag), as well as the linguistic accuracy of the names of concepts and explanations must be subordinated to it, but it also requires maximum flexibility regarding the ways in which a concept and all the connections surrounding the term can be accessed. Encyclopaedias that evolved from conventional, paper editions to computer-readable i.e. online editions often have within them the immanent structure of the "old" information organization. So no matter how many different new ways to search / find are offered to readers / users, they feel

that the links are not natural, that all the possibilities of interactivity and the logic of the functioning of the new media have not been completely used.

On the contrary, those lexicographic and encyclopaedia publications that started on the web are not "suffering" from distinct solutions not appropriate for the new medium. So as the greatest, inimitable success in bookselling was accomplished by a virtual bookstore that never operated in the real world (Amazon), and in sales through classified ads Craiglist which also did not have a paper start, among encyclopaedias by far the greatest importance has had the Wikipedia, the encyclopaedia that in only 10 years (appeared in 2001) grew to over 12 million articles in 279 languages [3]. No need to remind that only an online edition exists – if Wikipedia would be published in books the size of 25 cm and with thickness of 5 cm (about 6 MB per volume), the English edition alone would have 750 volumes.

Another question that arises when we analyze the representation of people or concepts is: what are the criteria upon which it is decided whether a term will be included? When after five years the figure of one million articles was exceeded, people began wonder whether each and every possible person, geographic concept, event or idea belongs in the encyclopaedia. If the intention was to cover every aspect of human knowledge then we have to raise the question of triviality of some concepts. If more stringent editorial policies are introduced to ensure the reputation of a source that has credibility, there remains still the question of inclusion, what has been left out and why? [4]

There is a lot of discussion about this problem both among the authors and editors of Wikipedia, as well as in the academic and professional community. The question that hovers in the air can be reduced to this: Could a possible criterion for the "publication" or "remaining" of a concept or term in the Wikipedia be his popularity, i.e. the number of people who using a search engine look for exactly that term? If this were the sole or even only the most important criterion, then Wikipedia would soon turn - at least by content - into a "yellow" publication, with all the negative connotations of that colour. It is clear that the inclusion of a concept that will never be sought burdens the encyclopaedia. Besides in every country (or in every language) there must be thousands of those who would like to prepare a text about themselves or a favourite character, theme, or the like, not thinking at all whether anyone will ever (other than themselves) ask the question to which such a text could serve as an answer.

---

[3] http://www.wikipedia.org/

[4] Hellweg, E: *The Wikipedia War; A recent high-profile dispute over the user-written encyclopedia's veracity has the site rethinking some of its rules*. MIT Technology review, Friday, December 16, 2005

3 Wiki – how much do we know each other

As we speak today about the "googlization" of information (what is not on Google may as well not have happened), the "powerpointization" of the presentation of facts or opinions (because the structure of the presentation adapts to the logic of concise entries), or the "facebookization" of personality (if you do not have a profile on that social network it is as if you do not even exist), thus we can safely say that today we encounter the "wikization" of education. There is almost no student who - to prepare a seminar, an essay or just home-work - will not at least in part use the definitions, explanations and/or references listed in this online encyclopaedia. The cause is - apart from an already formed habit, not to say addiction - also the fact that the text from Wikipedia, in the vast majority of cases will be found on the first page of the search results dis-play, regardless of the browser used. If the knowledge mentioned there is used only as a starting point for further research, this in itself not a bad thing, but mostly it is simply taking over finished units, including illustrations and other supplemental material.

In this study, we centered around the quality of "coverage" (taking into account only the formal elements, i.e. the number of lines) of some close to us and we might say self-explanatory lexical units in different Wikipedias. We conducted an analysis of the representation of prominent individuals in the above-men-tioned four languages or editions (whatever you want to call it), and four world languages (English/EN, French/FR, German (NJ) and Spanish (SP)). We ob-served the presence of writers and athletes in those languages. At the very be-ginning we chose the subjects: seven prominent members of the Department of Literature of HAZU (Croatian Academy of Arts and Sciences), and 5 well-known writers who are not members of the Academy. We took into account only living persons, active (or at least recently active) individuals. Previous testing showed that these are the most often represented people in the Wikipedia (e.g., scientists or artists are far less common). What was the case?

Table 1

Members of HAZU – Department of Literature	(HR)	(BS)	(SR)	(SH)	(EN)	(FR)	(NJ)	(SP)
Ivan Aralica	114	35			55			
Nedjeljko Fabrio	40							
Ivan Kušan	40	15			15			
Slobodan Novak	55							
Luko Paljetak	80							
Pavao Pavličić	75	80	60		5	6	115	
Goran Tribuson	70			35	140			

Table 2

Writers who are not members of HAZU	(HR)	(BS)	(SR)	(SH)	(EN)	(FR)	(NJ)	(SP)
Slavenka Drakulić		50		42	55		30	
Miro Gavran	50			50		5	70	
Miljenko Jergović	31	55			7	16	47	
Predrag Matvejević	80	55			45	75		
Dubravka Ugrešić	80	75	75	75	55		70	

First, we analyzed members of the HAZU: Ivan Aralica, and Ivan Kušan are represented in the Croatian, Bosnian and English edition; Nedjeljko Fabrio, Luko Paljetak and Slobodan Novak only in the Croatian; Pavao Pavličić is represented with articles in the Croatian, Bosnian, Serbian, English, French and German edition, and Goran Tribuson in the Croatian, Serbo-Croatian and English. For authors who are not members of the HAZU the situation is somewhat different, as we can see from the table.

The results are just calling for a comment! Not wanting to engage in comparisons or evaluations how legitimate one's representation or under-representation is, we can not avoid a few important conclusions. In the first place it is striking that none of the authors has an article in all the languages. Pavao Pavličić and Dubravka Ugrešić have articles in the most editions and they have the most lines: Ugrešić 430, Pavličić 330. At the same time it should be noted that Pavličić is the only one represented in all the world languages we took into account other than Spanish. On the other hand, only Slavenka Drakulić has no article in the Croatian edition, and two authors, Slobodan Novak and Luko Paljetak have an article only in the Croatian edition. Looking at the languages particularly, it is evident that our authors virtually do not exist in the French edition, and the situation is only slightly better with Germany. At the same time one fact should be taken into consideration; the German edition covers parts of Switzerland and Austria, where - or it is at least presumed so - the cultural connections, not only because of a shared history, are far greater than in the case of other countries and languages. On the other hand, the English edition is by far the largest, and was first launched, so in part this may account for the fact that it covers by far the greatest number of authors - as many as eight. Finally one should also say that in Spain Croatian writers do not exist - if we take as a criterion their representation in the Spanish edition of Wikipedia.

Concerning how familiar we are with each other in the territory of former Yugoslavia, seen in relation to the languages we have no problem communicating with, and not even mentioning the common history, it is absolutely startling that out of the 12 monitored authors only 2 are represented in the Serbian edition, 4 in Serbo-Croatian (?), and 7 in the Bosnian edition. The evaluation of who is relevant for the culture and education in a particular environment in the case of Wikipedia, of course, is not made by a ministry, commission or any administra-

tive authority, but by an assumed interest of those who write and edit the articles, as well as those who appear as users of these publications.

Of course, the analysis of individual qualifications that are given to different authors in different editions would take us far, so for example, Dubravka Ugresic is a Croatian (HR), (EN), (NJ) or Yugoslav, Croatian, Dutch and international (BO), (SR) (SH) writer. Predrag Matvejevic is a Bosnian and Croatian (HR) and Yugoslavian (FR) author, while the other editions (wisely?) do not take sides in this respect.

There is most information on Miljenko Jergović. It is the result of the popularity of his works which were translated in a number of foreign languages. He is also the only writer we included in both countries; considering his long-term work in Bosnia and Herzegovina, and the last 15 years of living in Croatia. Aleksandar Hemon is another writer for whom there is also plenty of information. He is not only a writer but also works in the film industry, he lives and works in Chicago, so this is another reason for a relatively large number of lines. Abdulah Sidran is mentioned in four Wikipedias (HR, BiH, EN and FR), which is probably the result of his numerous awards for screenplays and published works in many countries. In the Serbian Wikipedia there is only information on Hemon, and the German Wikipedia has only information on Miljenko Jergović whose many works were translated to that language.

The writer Dobrica Čosić has the largest number of lines, 421 respectively, which is probably the result of his great political involvement. From the 4 world editions of Wikipedia, the one in English has the largest number of Serbian writers represented - 5 out of 7, while the Spanish edition has the least number. The Bosnian edition offered a surprisingly small number of results, from the 6 authors offered there are articles only on two writers. One of them is Milorad Pavić, who has texts in most of the Wikipedias. But although all four editions of the world Wikipedias write about him, the Serbian, Serbo-Croatian, and even the Bosnian edition, the Croatian edition of Wikipedia does not have a single line on him. It should be noted that Milorad Pavić was still alive at the time of this research (he died on November 30th 2009 and the survey was conducted during the summer of 2009). Dragan Velikić is mentioned in the Serbian Wikipedia, then in the Croatian, because he was born in Pula, as well as in the German, because he was the ambassador of Serbia in Vienna.

Besides writers some athletes have been selected, all very well known, starting from the fact that the boundaries of individual cultures or languages are much easier crossed in a world where sports and what we commonly call "show business" have long since become globalized. This proved to be correct.

Representation of athletes was as expected far better, almost all of them appearing in the selected world languages, with an absolute record in the representation of Goran Ivanišević, with ~ 1450 rows, Janica Kostelić with ~ 1030, Blanka Vlašić with ~ 740, Toni Kukoč ~ 700, and Davor Šuker with ~ 560 lines. Janica Kostelić and Blanka Vlašić are the only ones represented in all the

306

languages/editions, and Goran Ivanišević still has far greater significance to other (international) languages, than in his own region. It is simply unbelievable that there is not even a word on Ivanisevic in two editions (Serbo-Croatian and Bosnian), but the number of lines dedicated to him in the English edition (by far the biggest article!) only confirms that the glory of winning Wimbledon is slow to fade. Toni Kukoč and Ivano Balić are in every edition except Serbo-Croatian while Davor Šuker does not exist in the Serbian and Serbo-Croatian edition.

The Bosnian websites write relatively little about the athletes of their own country. Regarding the "non-football" athletes there are only one karate athlete, and a chess player who have accomplished internationally relevant sports results. Bosnian football players are playing for German and Italian clubs, so they are mentioned in almost all the Wikipedias. Exceptions are the "Serbo-Croatian", which did not offer in its results any of the Bosnian athletes we have included, and the Serbian, which has a few words only on Džeko and Salihamidzic. The Spanish Wikipedia has completely forgotten about Edin Džeko, but the English and German Wikipedia have offered the most extensive information on Bosnian athletes.

In all of the Wikipedias there is the most information on Novak Đoković. It is interesting that some of the foreign Wikipedias contain more information than the Serbian one. The Bosnian Wikipedia has information only on one famous athlete from their neighboring country. The Spanish Wikipedia dedicated the most lines to Ana Ivanović, as a result of her former relationship with a famous person from that country. Jasna Šekarić was declared the world's best shooter of the 20th century by the World Shooting Federation, and she is mentioned in only four of the Wikipedias. The English Wikipedia has the largest number of lines on Nemanja Vidić because he plays for Manchester United. Milorad Čavić is one of the world's most famous swimmers, but by the number of lines in the Wikipedias this might not be concluded.

The purpose of this analysis naturally was not to lament over how much others know about us because then we would then have to start with ourselves first and analyze how much we know about others, and whether it is enough. Instead several questions should be asked: Is it more important to write "for ourselves", that is, to work on creating the best possible encyclopaedia designed for the audience (mostly students who are the most frequent users, but also others) in Croatia and for all who use the Croatian language, or should we systematically work on the presentation of "our" issues in the editions in other languages; whether it be persons or texts from the sphere of politics, history, or (most widely understood) culture and art. In the globalized world of today, is it more important to have all the answers "at hand" and who or what will be written about us left to the good will of the authors in the world (and perhaps some of the more ambitious ones at home)? Should someone be taking care about this?

In the endless debate about the quality and accuracy of what is published in the Wikipedia (the bibliography of articles on this topic has already reached a few

hundred units)[5] a new factor recently appeared: the more relevant the concept is i.e. the greater the interest for it is the greater the chance is that it will be corrected i.e. that a large number of readers/reviewers/authors will not allow inaccurate or biased information. Accordingly, if there are mistakes – they are in those places that are not much visited. The quality of the texts is thus proportional to the interest, and the number of reviewers (reader response) is what ensures accuracy.

Finally we return to the issues raised in the debate on the possibilities of finding relevant information by search engines. If we agree that we will get relatively successful answers to questions starting with who, what, where, when, and much less successful to those beginning with how or why, then this answers the question on the role of the online encyclopaedia in the transfer of knowledge and can even suggest an answer to the ever present dilemma: what to give for free, and what to charge (either by a subscription system, or as a single answer)? The value of the knowledge contained in the answer to the question posed with how or why is "value added" to the encyclopaedia; it is what distinguishes it from a dictionary or a search engine.

All these questions are definitely matter for further research; therefore this paper can be seen only as a foundation review of the topic intended to serve as a basis for research.

References

Berners-Lee, Tim: Weaving The Web: The Past, Present And Future Of The World Wide Web. Orion, London, 1999.

Britannica reaches out to the web. http://news.bbc.co.uk/1/hi/technology/7846986.stm

Cohen, Noah: Wikipedia Tries Approval System to Reduce Vandalism on Pages. http://bits.blogs.nytimes.com/2008/07/17/wikipedia-tries-approval-system-to-reduce-vandalism-on-pages/

Giles, Jim: Jimmy Wales' Wikipedia comes close to Britannica in terms of the accuracy of its science entries, a Nature investigation finds. Nature 438, 900-901 (15 December 2005).

Hefferman, V: Lexicographical Longing. May 11, 2008, http://www.nytimes.com/2008/05/11/magazine/11wwln-medium-t.html?_r=1

Kolbitsch, J. Et. Al: Dynamic Adaptation of Content and Structure in Electronic Encyclopaedias Texas Digital Library, Vol 8, No 3 (2007), http://journals.tdl.org/jodi/issue/view/35

Kung, Lucy; Picard, Robert; Towse, Ruth: The Internet and the Mass Media. Sage, London, 2008.

Lee Rainie; Tancer, Bill: 36% of online American adults consult Wikipedia. Pew Internet & American Life Project, April 2007

Manovich. Lev: The Language of New Media. The MIT Press, Cambridge, 2001.

Tapscott, Don: Growing Up Digital – How the Net Generation is Changing Your World. McGraw-Hill, New York, 2008.

The battle for Wikipedia's soul. The Economist, Mar 6th 2008. http://www.economist.com/printedition/displaystory.cfm?STORY_ID=10789354

Viégas, F. B. et. al: The Hidden Order of Wikipedia. http://www.research.ibm.com/visual/papers/hidden_order_wikipedia.pdf

---

[5] Giles, Jim: *Jimmy Wales' Wikipedia comes close to Britannica in terms of the accuracy of its science entries, a Nature investigation finds*. Nature 438, 900-901 (15 December 2005)**.**

Communication in an Electronic Society

Maja Živko
Department of Information Sciences
Faculty of Humanities and Social Sciences University of Zagreb
Ivana Lučića 3, 10000 Zagreb, Croatia
mazivko@ffzg.hr

Summary

Communication is a form of interaction in which messages are sent through channels by using a contextual code. The electronic revolution has brought mass media communication into life. Computer-based communication is the first mass communication medium in which there is a real interaction between the sender and the receiver. This communication has encouraged progress in social, economic, political, cultural, and educational sectors of human life. This article discusses the positive consequences of the expansion of the online world. E-based forms of communication bring integrity to people's lives through online chatting, discussion boards, and e-mails, lead to economic prosperity through marketing and advertising, help to strengthen democracy, encourage cultural globalization and expand knowledge through a large number of educational resources available online.

Key words: mass media communication, electronic communication, feedback, economic and political progress, network, time and space in virtual reality

Introduction

Communication is form of social interaction that involves exchange of information. It engages a sender and a recipient in a dynamic activity of communication. Over centuries, communication has undergone some major development and at least three "revolutions". The first revolution in communication began with the invention of writing systems - the script. However, originally, writing and reading, i.e. access to codified knowledge, remained a privileged activity of the literates and the educated stratum, whereas the experience of the common men remained limited to the face-to-face oral culture. The second revolution came with the invention of mechanised printing, which ultimately made mass literacy possible. This so-called "Gutenberg revolution" also enabled mass education to all social classes. Written text has dominated the structure of knowledge, communication, and culture. The latest third revolution has involved the invention of computer-based information that has given individuals

the power to communicate with other through text, graphic images, and sound instantly and globally.

The impacts of this third revolution are also the topic of this paper. It tries to answer two main questions. What has been the main impact of electronic media on social, economic, political, and educational progress? In what ways do electronic media and Internet-based channels of communication change our worldviews and ways in which we interact?

Computer-mediated communication

The information revolution has created the so-called "information society". Mass media communication started with the electronic revolution as a new way of communication by the use of electronic signal transmission. Through an electronic signal, communication is the process whereby a signal is transmitted from a source via various channels to a recipient. First, there is the transmission and then decoding of signals. The signal is sent from a beginning point through a channel such as a wire or a cable to a receiver. The amount of information is coded and defined by the type of an electronic signal. In e-society, communication is realized by using a sender-channel-receiver pattern. These screen-based multimedia communication networks give us the ability to communicate worldwide without the limitation of time and space. By the use of the Internet, people can transmit and receive large amounts of information quickly to and from individuals and groups from around the world.

Culture of real virtuality

The culture of computer mediated communication is usually referred to as the "culture of real virtuality" (Castells 2003). This is because the virtual reality effectively becomes a real virtuality. In this reality we live reality via electronic media in which there is no real time and no real space. This communication should ensure full participation in a multilingual information society by all citizens, strengthen science and technology, and stimulate international liaison and cooperation.

Computer media and economic prosperity

Computer media is conducive to economic prosperity. Many economic activities previously oriented to the local markets are now expanding their horizons to the global level. The development of information networks and the acceleration of technical progress are leading to an increase in internationally oriented services. Many organizations use electronic communication facilities such as the World Wide Web to enhance their teamwork because many individuals can simultaneously work on the same documents, hold meetings and integrate results of research at different locations. Moreover, the accumulation of capital is also enabled by networks. Capital can be invested globally in all activities via

these networks such as the IT industry, media business, agricultural production, health, education, technology, transport, trade, tourism, arts and culture, property, religion, entertainment, and sport.

Advertising and sponsorship online
Advertising is creating mass markets. Powerful forces use e-based media to influence us with their ideologies for their commercial purpose. Advertisers and other marketing actors use bulletin boards for trade and commerce. According to Vivian (2010:3), "the most obvious of the media messages designed to persuade is advertising. Organizations interact with their various constituents differently - employees, board members, customers, partners and others by using numerous media - text, graphics sound, video, etc. - into a single message."Another important aspect of the Internet-based sphere is sponsorship by businesses and organizations and support for discussion groups about issues and products that are relevant to these specific organizations. For example, sponsorship gives an opportunity for further communications and marketing.

Informatics and the global economy
In electronic society economy is electronic and global. The productivity and competitiveness of economic factors depend on their ability to effectively create, process and apply the information based on knowledge. The core of production, consumption, and circulation as well of their components (capital, manpower, raw materials, management, information, market technology) is globally organized. The productivity and competitiveness is created via a global network of interaction. Economically there is an increase of productivity and profitability as sources of wealth for any nation. In order to increase profit, expand the market, and accelerate the turnover of capital, productivity must be increased. This requires good communication skills which in turn is crucial for any success of information technologies. it is necessary to increase productivity, to expand the market and to accelerate the turnover of capital. It requires communication skills and, in turn, that enables information technology.

Capital network and computer network
A "networked society" is a society, in which, for the first time in history the basic unit of the economic organization is not the individual, the family, an enterprise or a state, but a network. A network consists of the connection of nodes and centres between multinational companies. Castells coined a new term - *metanetwork* - to describe this network of financial capital (2000:18). Capital metanetwork is intertwined with the computer-based network. Computer-based networks take many forms in today's organizations, including personal contact networks, flows of information within and between groups, strategic alliances among firms and global network organizations. It allows geographically distrib-

uted groups to communicate interactively and simultaneously and it leads to economic progress.[1]

Democracy in e-society

The first public communication was only a one-way communication and the authorities controlled information. Mass media electronic communication changed one-way communication into a democratic two-way dialogue. In the informational-communicational revolution there is a reversible form of mass communication. In e-society the computer-based media are platforms that mediate different opinions, views, theories, options, global-projects, investigations, economics, science, tourism, marketing, but also cultural and political views. The whole community should have access to the technology of electronic networks, and be trained in how to make the best out of them, so that they have an equal opportunity to fully participate in the economic and social benefits of the information revolution.

Participartory-democratic model in e-society

Computer-based media enable autonomy (from political, economic, and military restrictions), decentralization of communications systems and openness to global communications. In e-society everyone can function independently through social interactions. Decentralization of communication system disables imposing views and opinions of a group of people and other citizens on the whole society. In contrast, computer-mediated communication allows an active communication position for all participants in the communication process, not only political leaders. The citizen in e-society is not only a recipient but also a creator of information. In such a participatory-democratic model of e-society everyone can participate in the same way.

One voice in e-society

A truly democratic society in the theoretical sense of democracy can only be established if the voice of each individual becomes part of a larger project that allows empowerment of all. Each person's voice is valuable and essential for the "critical diversity" that is needed to raise important questions, notice oppressive power relations, criticize social conditions, and reach new levels of consensus and understanding. Individuals can state opinions and request feedback to grow in social understanding. They can participate in moderated discussions in which everyone in the "audience" can "speak" and contribute to the public discussion. School, business people, professionals, workers, and researchers can conduct projects with groups from across the world. They can share their interests and problems and generate new ideas as a global team. Democracy requires responsible individuals who are prepared to subject their be-

---

[1] Castells also calls the new way of capitalism as *informationalism* (2000:52).

liefs to public criticism and participate in the collective control of social life. All people with an online access have the power to enter a new social space for dialogue.

Inequality in e-society

New e-based technologies can help in employment, development of jobs, and improve job quality in some countries. They also foster an increase in political awareness, participatory democracy, mutual tolerance, and enable a more open, peaceful dialogue between the public and the elites. Today's information society brings changes, which have both benefits and disadvantages. E-society is strong and globalizing, but some parts of the population that are not privileged are excluded from the benefits it brings. People who do not have access to computers, knowledge of the Internet, and other related technological competence miss out on these opportunities to actively participate in the transformation of culture. They are not given the freedom, equality, justice, democracy, and autonomy that the Internet offers.[2] Lengel argues that, "disempowered people in disempowered places like the poor often think their voices are not adequately heard..." (2004:200).

Education by computer media

The Internet is one of our major information sources. Through computer-based media we learn almost everything we know about the world beyond our own immediate social environment. Computers have a large memory capacity and can mediate communication that "computerizes", i.e. provides in an electronic form, all selected human knowledge. This makes the work of researchers more time-effective, less costly, and allows them to invest their energy in new inventions. Less energy and resources is used on gaining old information and more energy is spent on discovering new information. On the World Wide Web the global electronic network, there are books, magazines, newspapers, and many other knowledge sources. Computers facilitate the communication between teachers, students, and parents. Students can now be tutored online by teachers from universities or schools in different countries. The Internet can be used to provide schoolwork, training courses, and other educational purposes. Chatting is also a good way to learn because students can ask questions online about things they do not know. It has been proven that computer-assisted instruction produces more learning than conventional classroom instruction.

Social globalization by networks

The mass media binds communities together by electronic correspondence and chatting. Entertainment online such as the movie, music, games, news, and

---

[2] There is a marginalization of those who are illiterate in information technologies.

magazine industries also connect people by exchanging their opinions and sharing experiences. In brief, the Internet enables users to participate in a global dialogue. To cite Rehm (2004), "perhaps the greatest benefit of the Internet is the opportunity to create a widespread interchange of ideas and build far-reaching relationships among diverse people over the entire world." Electronic communication allows discourse with a large number of correspondents, over wide geographical areas, with no more effort or cost than is required to send a single message locally.

Obstacles to e-social globalization by networks

Misinterpretations

In e-society there is no voice inflection and body language so misunderstandings and conflict can occur. Subtle humour tends to get lost in electronic communication which can lead to confusion or misinterpretation. People have to make sure that the others are able to detect humour or irony. For these reasons, the e-community has developed symbols called "emotions" to indicate humour, feelings, or other grimaces. However, even if we use these, it is always safer to remind others that you are being ironic or funny. The wrong interpretation of messages can often generate unexpected angry responses called *flames*.

Redefining the community

The nature of public interests and ideas are vast and often conflicting. Because the public e-sphere is virtual it offers multiple representations of a plurality of world. The Internet simply does not have the capacity to facilitate or even allow stable societal norms.

Conclusion

Electronic communication was invented together with computers. It has been spread all across the world by the creation of the so-called World Wide Web. WWW is a global network of networks that enable the flow of messages among communicators across time and space. Computer-based communication allows for a new form of communication, which can lead to an economic and political progress, social globalization, and even improvements in education. In e-society individuals interact, exchange ideas, learn, provide social support, conduct business, play games, and even participate in politics. The problem is that information systems are not equally distributed across society and some people have disadvantaged access to computers. In other words, also the e-society leads to social inequalities that include networked companies with global manpower and those of the so-called excluded population. Therefore, the question remains how to achieve a worldwide and fair access to computer-mediated communication so that every country, group, and section of the population would be fully integrated this new electronic world. It is up to all of us to try to find solutions to this problem.

References

Baldwin, E. E. Transformative professional practice: Overcoming ambivalence, building community. *//Kappa Omicron Nu FORUM.* 8 (1995)2; 21-37.

Castells, M. Uspon umreženog društva. Zagreb: Golden Marketing, 2000Castells, M. Moć identiteta. Zagreb: Golden Marketing, 2002

Castells, M. Kraj tisućljeća. Zagreb: Golden Marketing, 2003

Giroux, H. A. Literacy and the pedagogy of voice and political empowerment. //Educational Theory, 38 (1988)1, 61-75

Helgesen, S. The web of inclusion: A new architecture for building great organizations. New York: Doubleday, 1995

Hoffmann, Ch. Language, culture and communication in contemporary Europe. Great Britain: WBC Book manifactures, 1996

Lockhard, J. Progressive politics, electronic individualism and the myth of virtual community. *//Internet culture*. New York: Roudedge, 1997, 219-231

McLaren, P. Critical pedagogy: Constructing an arch of social dreaming and a doorway to hope. Journal of Education, 173 (1991)1, 9-34.

Melanie J. Norton, Introductory Concepts in Information Sciense. USA: Thomas H. Hogan, 2008

Monge, P. R., Contractor, N., Theories of Communication Networks. Great Britain: Oxford University Press, 2003

Plenković, M. Komunikologija masovnih medija. Zagreb: Barbat, 1993

Rehm, Marscha, L. The Internet as a Practical Problem: Empowerment in the Electronic Global Village. 2004. http://www.kon.org/archives/forum/11-1/rehm.html (2004-08-26)

Thurlow, C. C., Lengel, L., Tomic, A. Computer Mediated Communication. USA: Sage Publications, 2004

Vivian, J. Media of Mass Communication, 2010

New Media – Ethical Issues

Mila Nadrljanski
Faculty of Maritime Studies – Split
Zrinsko Frankopanska 37, Split, Croatia
milamika60@yahoo.com, mila@pfst.hr

Đorđe Nadrljanski
Alfa University Belgrade
Vase Staića 6, Belgrade, Serbia
djordjen44@gmail.com

Zita Buzasi
Alfa University Belgrade
Vase Staića 6, Belgrade, Serbia
zizy84@yahoo.com

Summary

The aim of this paper is to examine the phenomenon of ethics in the media. Its purpose is to present problems in the field of media ethics, which could possibly increase the critical awareness of the effects of this problem. The tasks of this study are: to identify the media influence on behaviour, to find examples of common ethical dilemmas and typical problems in the field of media ethics, to examine people's opinions related to a given issue, to identify similarities and differences between employees (referring to journalists and editors) and the unemployed in the media and to examine the ethics of the four different media (television, press, radio and Internet). The methodology of the study is research, personal observations and discussions, as well as interpretations of the phenomenon by various authors who have written about it. The practical significance of this work is to stimulate thinking about the theme as well as to provide some information that might be useful in any subsequent research.

Key words: media, media research, ethics

Introduction

The aim and purpose of this paper is to examine problems in the field of media ethics, in order to possibly increase the critical awareness of the effects of this problem. The tasks of this study are: to determine what are the influences of media on behaviour, to find examples of most common ethical dilemmas and typical problems in the field of media ethics, to examine people's opinions concerning these issues, to identify similarities and differences between media em-

ployees (referring to journalists and editors) and media consumers, as well as to examine ethics of four different media (television, press, radio and Internet). Methodology of this study is research, personal observations and discussions, as well as various authors' interpretations of this phenomenon. The practical significance of this paper is to stimulate discussion about the topic and provide some information that could be useful in future research.

Could ethics be learned?

There are two answers to this question. Some cynics claim that ethics cannot be a subject of a study because it raises questions which cannot be answered clearly. Besides, some skeptics claim that being familiar with ethical principles and norms does not automatically make a person moral. Another school of thought, which is represented by optimistic supporters of formal ethics education, think that ethics is a subject such as mathematics, physics or history, with its set of problems and unique methods of their solution (Day, 2008). A psychologist James Rest made an attempt to summarize scientists' attitudes: People between 20 and 30 experience dramatic changes in thinking and resolution of ethical dilemmas. These changes are associated with understanding of society by the individual and the individual's role in society. The degree of changes in a person are directly associated with formal education in high school or university. Experience has shown that teaching ethics in schools is effective, i.e. that kind of education affects consciousness of individuals' morality. Moral attitudes and moral evaluation of a person have the most influence on their behaviour. Lawrence Kohlberg was especially engaged in problems of education and ethics. One of his first conclusions was that the individual's ability to deal with moral issues does not form suddenly. Just as there are stages of physical development of a person, there are stages in the process of moral development. In his book "Ethics in Media Communications: Cases and Controversies" Louis Alvin Day says that a lecturer teaching media ethics probably cannot teach moral behaviour in the sense that teaching ethical principles makes a morally impeccable person. Teaching ethics can promote moral behaviour providing means for making moral judgments, defending them, and then criticizing the consequences of one's choice. This process is known as moral reasoning. (Day, 2008).

On media

Media provide us with information about innovations and deficiencies for which we have no knowledge of our own as consumers. We meet the term "media" in everyday life and we are familiar with it. We use it without much thinking. Being ubiquitous, the phenomenon of media has been a habit for some time. Our age is therefore rightly labeled as "the era of media and information." But, beware! If you take a better look, you can see that the term of media can not be taken that simply. There is a large vagueness in use. Everything gets a popular label of media, but it is not always appropriate. Therefore, we should

318

consider the term of media itself. "Media are an integral part of communication. They are an important part of communication in society and among people. Media are used to make a communication successful, both in individual or mass communication. In communication, by the means of media, people convey messages, they use mime, gesture, letters, images, tone, in face to face, or printing or electronic transmission or storage techniques. Media are the means of transmitting messages. Media are mediating instances without which communication is not possible. The value of media is, therefore, in the effect that contributes to the success of communication. This effect is in mediation. It is important to clarify the phenomenon of communication. The word "media" comes from the Latin adjective *medium*, which means "the one in the middle", "medium". The application of the term is connected with "the idea of mediation, the idea of the holder of spiritual expression, which breaks loose from time and space dependencies. The concept of media in everyday life is used primarily as a collective term for technical means or instruments which serve for dissemination of information."(Nadrljanski, 2010, 51, 52). One of the basic requirements to define a certain term is a completeness of the definition, i.e. the possibility of its application in a broad subject and phenomenological field. When speaking about communication and the role of media in this primary social relationship, media cannot be given only the situational role of mediator, although this is their most important feature, and be merely something between the subjects of communication and serving in transmitting and exchanging messages. For the classification of media, two criteria are relevant within communicology framework. The first one is the way of perceiving mediated messages, by which we make a distinction between visual (all the press media), audio (radio) and audio-visual (film, television, Internet). Another criterion is a form of communication practice with the immanent medium, dividing media into: media for interpersonal communication (voice recorder, computer...), media for group communication (radio, television), media of mass communication (press, radio, film, television, theatre) and since the end of the last century, the medium of virtual communication (computer network).

The negative connotations of media are mainly associated with a creation of social culture of people who share and support the same interests and who are becoming more subject to media manipulations. The mass media are among the most influential institutions in a democratic society, at the intersection between citizens and their political, economic and social institutions. Besides, media workers are crucial for transmitting cultural values. They determine which values are important and they offer symbolic signs for standards of manner, including ethical behaviour. "Media are the primary source of information in democracy. Accurate and reliable information is the basis of a democratic process." (Radojković, Miletić, 2006, pp.95-96). "Young people are generally not familiar with the fact that the ownership structure can influence media in a large scale." (Miliša, Tolić, Vertošek, 2009, p.17).

On Ethics in media

"Using today's high digital technology, a journalist shapes reality and sells it as a product to users, ie public opinion, on the market. The ethics of this profession depends on self-control or self-censorship of a person who selects the product – a journalist. In the confusion and race for time of a modern man, and trying to follow some streams of modern communication in the new media and deregulation processes, there is another issue of responsibility imposed on a consumer of "media circus". Media should educate the public about the basic human rights. What is the objective of media quality, i.e. the method by which media achieves quality? Which is more important, ethics or a method of obtaining information? There is the key of the problem because ethics depends on the method. A good or bad method sets an ethical limit, within which is the ultimate ideal – the truth. The demands on media are like Sisyphean problem, but within the mass media limits they can be regulated. These demands are: freedom, quality and responsibility" (Panjeta, 2003, p.11). The media behaviour in this country and rules of conduct of professionals in it have always been imposed by someone else: government, ruling political parties, media tycoons close to them or other tycoons... Journalists were supposed to be completely adaptable: to follow the imposed rules strictly. The cunning ones, who had the specific talent to fulfill all the wishes of the powerful with a smile on their face – like the best prostitutes, could count on their mercy. Sometimes for the rich prize ... Everything else was less important: education, knowledge or talent. Effort and work. Especially facts." (Nikšić, 2004) Today's media professionals are more educated and trained than ever. Many of them come from universities unprepared to deal with ethical problems of the real world. It would be useful for students to face the difficult ethical dilemmas in the classroom, where they can discuss them rationally, without the pressure of harsh deadlines which await them in the future. Good professional ethics is something you should take care of and which contributes to the respect among colleagues. This is another argument in favor of a formal ethics education. According to Malović, media based on fiction began to increasingly influence the media based on facts, "a fictional story is much more interesting and exciting than a journalistic report" (Malović, 2007, 12). Miliša questions "How do media manipulations reveal social deformations?" (Miliša, Tolić, Vertovšek, 2009, p.19). Events are presented as an invented reality. "Most of the media content brings the social norms into question. Without stories about crime, violence, drugs and suicides, for example, newspapers as well as entertainment would be left without dramatic vitality. It would be unreasonable and unrealistic to delete all controversial content even if the effects on the audience are unpredictable. The goal should be devising strategies that would promote responsible treatment of antisocial behaviour in the media and to avoid approaches that encourage moral degeneration."(Day, 2008, p.320). Ethical issues related to the role of media in influencing antisocial behaviour include all three functions of mass media: news programme, entertainment and advertising.

To consider the information ethics issue means to stick only to the aspects related to the function that is given to journalists. These aspects are complex enough and it is completely reasonable to observe them from the standpoint of ethics. ***What does it mean to inform?*** It means to form the certain message with the intention of conveying it to someone. Information, as it is commonly understood, includes both content and form, as well as their transfer. Immediately after the media put the massacre at Columbine High School on the front pages, an incredibly large number of "plagiarisms" were reported across the country.

The question is how many people are media consumers in general. We ourselves are not aware of the fact how big the number is. It is an indicator of dependency and alienation. They do not perceive it as violence in the form of a reality show like Big Brother, The Moment of Truth, Wife Swap and various Mexican and Turkish soap operas etc. There is also a lot of rude behaviour in media which young people take as a normal form of behaviour. In the spring of 1994, Jerry Springer Show reached very low ratings, thus being in danger of canceling. Springer and his producer have brought the instant decision to turn the show into "a relationship show with lots of fights." The ratings increased by 100%. (People who were faced in the show started to discuss vigorously and often ended up fighting). Hate speech is more and more present. For example, the neo-Nazi group calling itself the SS Action Group placed a recruitment message on Warner Cable Channel in 1987. The message said: Join the American Nazis and break the red, black and Jewish power.

The truth in Media

We are witnesses of a large number of tabloids which publish a variety of lies about public figures, for the sake of exclusive story and good sales. The first article of the ethical code of electronic media, concerning the duties of journalists, states that reporting has to be based on the facts supported by evidence. Local media are more concerned with guessing than with the search for truth. Local scene media intrude into privacy of public figures and others, turning the entire reporting and creation of entertainment programme into voyeurism. According to Article 33, The Public Information Act of the Republic of Croatia, any private information or personal written record (letters, diaries, notes, digital recording and the like), a figure record (photograph, drawing, film, video, digital, etc.) and voice recording (tape, gramophone, digital, etc.) can not be published without the consent of the person to whom the information relates, or the person whose words, figure or voice it includes. The TV show "All for Love" by Emotion Production, broadcasted on TV Pink, often film and broadcast private conversations and emotional scenes without permission. The commercial for "Plazma biscuits" under the slogan "It won't let you grow up", influenced the thinking of several children who refused to take a biscuit with the excuse "I want to grow up, and Plasma would not let me." Also, the vast majority of TV hosts and announcers make serious mistakes in pronunciation, word choice and

stress and the population accepts such distorted expressions as standard ones. It is interesting to mention the quiz show "The Moment of Truth" which used to be broadcasted on Croatian TV Nova, in which the participant responds to intimate questions with YES and NO, while a polygraph measures frequencies. With each new level the participant wins more money, and the questions become more intimate and improper, offensive and appalling. When a participant gave a positive answer to questions like "Would you apply physical force to persuade someone to have sex?", "Have you had an intimate relationship with your friend's wife?" or "Have you ever given a child a piece of food from the floor?", the audience was not appalled, but applauded because "that brave and honest man" won a large amount of money. Also, some public figures are improperly dressed during daily entertainment programmes, and the same shows are imbued with ads and commercials with sexual content. The local press undoubtedly emphasizes female stereotypes. A woman is looking for a rich man, she cares more about her looks than education, she is always dressed up and visits "cultural" events. Any woman who does not recognize herself within aforementioned stereotype, in the media is presented as being "behind", far from prime time. Article 3 in The Code of Ethics for Electronic Media states that broadcasters must avoid stereotypes and biases when reporting about a community. They should oppose the interlocutors who express stereotypes and prejudices in interviews and discussions. Local media do not oppose the stereotypes and prejudices, they encourage them.

Research problem

The main scope of this study is examining ethics in the media (television, press, radio and the Internet). The aim of the study is to examine whether there are differences in opinion on ethics in media between media employees on the one hand and "media consumers" on the other hand. The aim is also to determine if there is a difference in opinion regarding the ethics on television, in the press and on the radio, as well as to examine the views on ethics in media generally. Research hypotheses are:

1. The assumption is that the respondents would give positive answers to questions generally related to ethics and morality.
2. Regarding the group of respondents who are media consumers, the assumption is that their attitude on the issue of ethics in media would be negative.
3. Another assumption is that the group of employees (journalists and editors) would have a generally positive attitude on the same issue.

The sample for this pilot[1] study consisted of 100 subjects (59 females and 41 males). The youngest respondent was 21 years old and the oldest one 62. The

---

[1] This study is a part of more extensive research project covering 700 respondents in Split Dalmatia County.

average age of the sample was 35. A half of the respondents (50 people) were employed in the media, and the other half were media consumers (not media employees).

The instrument used in this study was a questionnaire consisting of 87 items arranged into 4 groups: a group of items related to the general attitude of people towards the concept of ethics and morality; a group of items related to attitude towards ethics on television; a group of items related to attitude towards ethics in the press; and finally a group of items related to attitude towards ethics on the radio. The respondents had to circle one of three offered answers (YES, NO, MAYBE).

The main objective of this research is to examine the attitudes of adults regarding the following topics:

- Violation of privacy in order to discover the truth,
- Whether it is reasonable to publish so called "The Crime" column,
- Disgusting content on the Internet,
- Violations of moral standards and national security while providing information on the Internet,
- Economic interests related to the media,
- Explicit and disturbing images,
- Publication of promotional materials for cigarettes,
- Publication of promotional materials for alcohol,
- Unethical content on the Internet for children,
- Unethical facilities in media used in education,
- Unprofessional people as managers of various sites,
- Showing homosexual scenes,
- Showing explicit sex scenes,
- Hate speech,
- Showing pornography,
- Youth suicide,
- emotional injuries,
- Violence,
- Aesthetic values,
- Freedom of research,
- Influence on youth,
- Stereotypes,
- The racial and ethic minorities,
- Discrimination against women,
- Discrimination against people of different sexual orientation,
- Discrimination against older people,
- Discrimination against disabled persons,
- Prejudice,

- Ethics and morality in general (general questions).

The study was conducted in the town of Split, in November 2010. Respondents were told about the anonymity of the questionnaire. The study was conducted without difficulties. For each group of questions a particular hypothesis was set.

Research results

The first group of questions (ethics and morality in general)

As for the hypothesis related to the first group of questions (5 questions), the most responses obtained were positive, thus confirming the initial assumption. "Ethics is something that is taught, not something to be born with", was answered only 5% negatively and 82% positively. "Ethics in media depends on moral behaviour and attitudes of their creators (journalists and employees)", was answered only 7% negatively and 76% positively.

The second group of questions (ethics on television)

The hypothesis that most respondents would answer negatively regarding ethics on television was confirmed as the most negative responses were obtained. "I sometimes notice disgusting content on television," was answered 4% negatively and 92% positively. "Unprofessional people should not work on TV", was responded only 4% negatively and even 91% positively.

The third group of questions (ethics in the press)

The hypothesis that most respondents would answer negatively regarding ethics in the press was confirmed. "I sometimes notice disgusting content in the press," was answered only 4% negatively and 88% positively. "Unprofessional people should not work in the press", was answered 8% negatively and 84% positively.

The fourth group of questions (ethics on the radio)

The hypothesis that most respondents would answer positively regarding ethics on the radio was confirmed. "Unprofessional people should not work on the radio," was answered 90% positively and 5% negatively.

The difference in attitudes about media ethics between "consumers" and employed in the media

The hypothesis that respondents who are media consumers would have a negative attitude, and that employees would have a generally positive attitude on the issue of ethics in media has not proved correct because the responses of both groups generally coincide. The only significant differences are observed in "The press offend senior citizens," where 24 media employees responded negatively (48% of the group of employees), as well as 40 from the group of media consumers (80% of the group of consumers); "The radio is not trying hard enough to reduce stereotypes," where 3 media employees (6%) and 19 from the group of consumers (38%) responded negatively; "The radio encourages prejudice"

324

where 41 media employees (82%) and 24 from the group of consumers (48%) responded negatively; and "The radio has a huge impact on young audiences," where 45 media employees (90%) and only 18 from the group of consumers (36%) responded positively.

Conclusion

Interpretation of the phenomena of the same items in the studies of all three media can lead to a double conclusion. Namely, by confirmation of these assertions, it can be concluded that people believe that there are many incompetent people working in media. Or we can conclude that people are generally dissatisfied with the ethics in media and their specific content (which is not ethical), thus attributing those "errors" to people who work in the media, which forms the basis of their incompetence (we do not know whether this is a fact that people are acquainted with, or an assumption caused by dissatisfaction with the work of media and the supposition that incompetent people do what we expose to criticism). In this study of all three media items like "I sometimes notice disgusting content ..." and "Unprofessional people should not work ..." can be considered the way in which their non-concreteness may include a series of other items (which are concrete), therefore making these responses important.

The hypothesis which was not confirmed and which relates to the differences between attitudes of "media consumers" and media employees, claiming that employees would have a positive attitude on the issue of ethics in all three media, and that consumers would have a negative attitude, is based on a premise that consumers have a greater critical awareness of media ethics than those who create the media. The assumption is that political parties and advertisers dictate editorial policy in all media because media financially depend on parties and advertising production.

Since the examinees' responses in this study mainly depicted their negative attitude on the issue of ethics in media, we can undoubtedly conclude that media ethics in this country is interpreted as a major problem, and both media representers and media consumers are aware of it. We could say that the media life which we all live in, is ruled by anarchy instead of democracy, and that the codes and laws are obeyed and sanctioned rarely. On the one hand, media have "no mercy" in presenting and publishing anything that they stumble upon, persistently and successfully escaping from ethics, driven by the desire for wealth and fame. On the other hand, the "purist" media, in an effort to clean up the contents from the sin of immorality, make a tasteless creation without any "spicy" element, making it simply unpleasant. Thus, we are left to interpret, criticize, take measures that are in our power and to hope that what we think is a product of our intellectual and mental efforts, not served portion of the media pap.

References

Brigs, Adam; Kobli, Pol, *Introduction to Media Studies,* Belgrade, Clio, 2005.

Day, Louis Alvin., *Ethics in Media Communications: Cases and Controversies*, Belgrade: Medija centar: Plus, 2008.

Korni, Danijel, *Information Ethics*, Beograd, Clio, 1999.

Malović,S. (2007), Media and Society, Zagreb: University Library

Mediacentar Online (http://www.media.ba)

Miliša, Tolić, Vetrovšek (2009), The media and youth, Zagreb

Nadrljanski,M.(2010), Communication Science and Management, Split, Redak

Nedeljnik Vreme (http://www.vreme.com, Grujičić, Nebojša, an article „What is Džafer doing to-day?", 2002)

Nikšić, S., *Ethics in Journalism: A Handbook for Professional Journalists*. Belgrade: Centre for Media Professionalization, 2004.

Nova Srpska Politička Misao, a journal of theory and social research (http://www.nspm.org.yu)

Radojković, Miroljub; Miletić, Mirko, *Communication, Media and Society,* Novi Sad, Stylos, 2006.

The Centre for Media Professionalization (http://www.cpm.edu.yu)

The Code of Ethics of Electronic Media,

The Law on Public Information of the Republic of Croatia,

LANGUAGE TECHNOLOGIES AND LANGUAGE PROCESSING

Translation as a Business

Annette Hemera
Business Development Manager Europe
CSOFT International Ltd.
annette.hemera@csoftintl.com

CSOFT International Ltd.

As a leading provider of multilingual localization, testing and outsourced software development for the global market, CSOFT delivers quality and affordable language and technology solutions by combining advanced engineering capabilities, best-of-breed processes, professional project management and in-country linguistic resources around the world.

WW Headquarters Office
15th Floor, West Tower,
World Financial Centre
No.1 East 3rd Ring Road
Beijing, China 100020

The characteristics of a "good translator" have been repeated many times in many places. Advisors often explain how to translate a contract or operation manual; how to define adaptation; how to recognize a good translation memory (TM) and/or dictionary program; where to find reliable glossaries on the net and how to recognize problematic source files; to weight the number of words in your translation; and how to open a zip file. Consistently reliable quality and efficient use of tools are necessary but not sufficient to become a successful translator.

Unless these things are supported by thoughtful and consistent management, a translator cannot expect more than temporary success. While many translators live under the illusion that as freelancers the only thing they will have to do is translate, a good measure of true success is how well the translator can manage his or her one-person business.

Like it or not, in addition to translating, a freelance translator has to juggle all the tasks needed to keep a business running, no matter what the size. In this regard, there is no difference between Microsoft and John Smith Translations because a one-person venture has to traverse every single step in the "value chain" just like its corporate fellows. It has to establish the business, organize it, and maintain it at optimum operational level.

The freelance translator becomes the production manager, quality controller, operational director, supervisor, sales manager, advertising and PR specialist, IT

director, chief stockholder and board chair all in one. Sometimes the posts even conflict, and short-term interests can be very different.

Most translators dive head first into their companies without any training in the business world. Many go through the steps of getting a freelance operation permit, a computer and a stack of dictionaries, and proceed to bury themselves at their desks translating night and day. They treat anyone who offers accounting or tax advice, advertising recommendations, or chances to register themselves on a website at low cost as an impediment, and consider the effort of doing any of these things a pain in the neck.

However, a freelancer needs to spend a good portion of his or her working time planning, advertising, bidding for jobs, preparing invoices and accounts, and conducting cost calculations, to put a business on track to long-term success. Most translation schools focus exclusively on theoretical information and offer nil in the way of practical tips. So, professionals just entering the workforce have many difficulties to overcome that take away time and energy from the real job. Lack of knowledge about how to run a translation business, the outcome of intrinsic shortcomings in translation courses, is the basic reason for the huge vacuum on the Eastern and Central European language services market. On the one hand, here in Hungary we see a large number of theoretically trained but inexperienced translators hoping in vain to be entered into client databases, while on the other, we have clients unable to find translators to cover their day-to-day capacities fighting bitterly to secure good vendors. Since most specialized translators study translation to supplement a primary course of studies such as technology or law, they tend to treat a degree as a specialist translator as a lifebelt, a second profession to fall back on if they are unable to find a job in their primary field. If worse comes to worst, they can work as freelance translators, they think. Given the *ad-hoc* nature of the choice, most professionals who start businesses as translators have not surveyed the business climate, have not garnered basic information on the language services market, have not mapped out demands, and have no idea about minimum investment costs or, obviously, expected inflow. In other words, they have no idea what they are up against.

Evaluation of Free Online Machine Translations for Croatian-English and English-Croatian Language Pairs

Sanja Seljan
Department of Information Sciences,
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
sseljan@ffzg.hr

Marija Brkić
Department of Informatics, University of Rijeka
Omladinska 14, Rijeka, Croatia
mbrkic@uniri.hr

Vlasta Kučiš
Department of Translation Studies,
Faculty of Arts, University of Maribor
Koroška cesta 160, Maribor, Slovenia
asta.kucis@siol.net

Summary

This paper presents a study on the evaluation of texts from four domains (city description, law, football, monitors) translated from Croatian into English by four free online translation services (Google Translate, Stars21, InterTran and Translation Guide) and texts translated from English into Croatian by Google Translate. The aim of the paper is to conduct a machine translation evaluation of free translation services and to measure inter-rater agreement and the influence of error types on the criteria of fluency and adequacy. The evaluation is performed by students and the results are analyzed by Fleiss kappa and Pearson's correlation.

Key words: evaluation, free online translation services, English, Croatian, fluency, adequacy, error analysis, Fleiss, Pearson

Introduction

Although human-assisted machine translation (MT) still gains significantly better results than automatic translation (Bar-Hillel, 2003), the use of online translation tools has increased in recent years, even among less widely spoken languages. Despite the fact that it is desirable to have moderate to good quality

331

translations, there are cases where high quality translations are not of crucial importance (e.g. gist of a paragraph from a Web page, e-mail translation, basic information about a conference, a product, etc.). Evaluation from the user's perspective helps producers, but also examines the translation problems. However, most of the evaluation has been conducted only for widely spoken languages, which possess various language tools and resources.

The increase in the use of free online translation tools has caused an increased interest in the evaluation of these tools. The recent studies have raised questions regarding possible uses of online MT. Besides gisting, MT can be used in information retrieval, i.e. question-answering systems (Garcia-Santiago and Olvera-Lobo, 2010). As presented by this pilot research, free online MT tools are also used for homework translation, where, apart from pedagogical use, it is important to detect the inappropriate use (Hampshire and Porta Salvia, 2010).

In 1976 Systran launched its first MT system for the Commission of the European Communities. The first online free translation on the Internet appeared in 1997 by Babel Fish using Systran technology (Aiken et al, 2009). In 2007 Google Translate online translator appeared, relying more on the statistical approach and the comparison of matching probabilities, than on the rule-based approach. Ever since, it has been included in almost every evaluation study.

According to Kit and Wong (2008), misunderstanding of MT can be avoided by having realistic expectations and using appropriate text genre. Zervaki (2002) points out that in the case of simple sentences and SVO order, MT can produce acceptable terminology and syntax. However, in more complex sentences translations become incomprehensible. Online translation services mostly serve popular languages and there is a considerable difference in the quality of translation dependent on the language pair and the type of text being translated.

MT evaluation has been used not only for evaluation of different commercial or online systems, but also during system development.

This pilot research performs evaluation of free online translation services for less widely spoken languages, such as Croatian, and measures inter-rater agreement. The human evaluation, based on the criteria of fluency and adequacy, is enriched by error analysis, in order to examine the influence of error types on fluency and adequacy and to use it in further research.

Due to a small test set, the results should be taken as preliminary. After the related work section, the test set and evaluation procedure description is given. In addition to the evaluation results, error analysis is given and Pearson's correlation and Fleiss kappa results are presented.

Related work

DARPA (Defense Advanced Research Projects Agency) has presented the outcome of the research on the evaluation of various MT systems, based on the black box methodology, as presented in White and O'Connell (1996). In order to avoid the subjectivity of human judgments, it has been suggested to

decompose judgments into adequacy, informativeness and fluency (White et al, 1995).

The study described in Garcia-Santiago and Olvera-Lobo (2010) analyzes the effectiveness of the translations from German and French into Spanish obtained through Google Translate, ProMT and WorldLingo. They emphasize the benefits that would be gained when research studies would use the same scale of human assessment.

In the study performed by Dis Brandt (2011) three popular online tools have been evaluated (Google Translate, Tungutorg, Apertium). Despite the fact that the number of errors significantly decreases after human editing, some tools show significantly better results.

In the study conducted by Aiken et al. (2006) random Spanish sentences from two introductory Spanish textbooks and two web sites have been translated into English by Systran and evaluated by undergraduate students.

In the study presented by Kit and Wong (2008), Babel Fish, Google Translate, ProMT, SDL free translator, Systran and WorldLingo have been evaluated using BLEU and NIST for translating from 13 languages into English. Systran and Babel Fish have proved to be the best for the majority of language pairs, while Google Translate has proved the best for translation from Arabic and Chinese into English. ProMT has proved better than the rest for Portuguese-English and WorldLingo for Swedish-English language pair.

The evaluation of MT systems is of high importance in MT research and product design. The evaluation is done either to measure system performance (Giménez and Màrquez, 2010; Lavie and Agarwal, 2007) or to identify weak points and/or to adjust parameter settings of different MT systems or of a single system through different phases (Denkowski and Lavie, 2010a; Agarwal and Lavie, 2008).

The evaluation of MT is done using different language independent algorithms, mostly BLEU (Papineni et al., 2002) or NIST (Doddington, 2002). In order to obtain metrics that give results closer to human evaluation results, there is a need for qualitative evaluation of different linguistic phenomena integrated with statistical approaches (Monti et al., 2011). In the study presented by (Denkowski and Lavie, 2010b) the factors that constitute "good" or "bad" translations are discussed, with the focus on difficult points of inter-rater agreement.

Experimental study

The evaluation of free online translation services has been performed from the user's perspective, by undergraduate and graduate students of languages, linguistics and information sciences, who have attended one or more courses on language technologies at the University of Zagreb, Faculty of Humanities and Social Sciences. The texts from four domains (city description, law, football,

monitors), have been translated from Croatian into English and from English into Croatian.

Translations from Croatian into English have been obtained from four Internet translation services with Croatian language support:

Google Translate (GT) - http://translate.google.com

Stars21 (S21) - http://stars21.com/translator

InterTran (IT) - http://transdict.com/translators/intertran.html

Translation Guide (TG) - http://www.translation-guide.com

GT is a web-based translation service provided by Google Inc. It is a statistical MT based on huge amount of corpora. It currently supports 57 languages. The Croatian language has been supported since 2008. Although S21 service is powered by GT, the translations are not always the same, probably due to different pre- or postprocessing techniques. Translation Experts Ltd. Company provides MT services using IT, which is powered by NeuroTran and WordTran, which translate sentence-by-sentence and word-by-word, respectively. NeuroTran is a hybrid system that uses a combination of linguistic rules, statistical methods and neural networks, as well as text analysis, to determine the context for the unique type of lexical selection. Although TG is powered by IT, the resulting translations differ. While Croatian-English translations have been obtained from four mentioned services, the English-Croatian translations have been obtained only from GT.

Test set description

The students have evaluated four illustrative excerpts translated from Croatian into English from the domains of city description, law, football and monitors, with 9, 9, 7, 9 sentences, respectively.

Table 1. Text statistics

		Total no. of words	Min sentence length	Max sentence length	Avg sentence length
En-Cro	city	160	11	27	17.7
	law	227	5	43	25.2
	football	160	3	32	17.7
	monitors	112	3	28	12.4
Cro-En	city	119	5	22	13.2
	law	196	9	46	21.7
	football	133	9	29	19.0
	monitors	123	3	22	15.4

Table 1 presents the average sentence length per domain and translation direction. The text about the city and the legal text are the same for both directions (indicating that English texts are on average 20-25% longer), while

the text on football and the text on LCD monitor differ from one direction to another.

Evaluation

The evaluation of Croatian-English (Cro-En) translations has been performed by 48 students (64.6% undergraduate, 35.4% graduate), who are either finalizing their studies of the English language and information sciences, or who have been learning English for about 10-12 years. 75% of the students have attended one or more language technology courses, such as Machine Translation, Translation Memories as Translation Tools or Computer-Assisted Language Learning. 83% of the students have declared to have had previous experience in translating, out of which 20% in professional translation, 52.5% for the faculty needs, and 27.5% for personal needs.

According to the answers, 39.6% of students use online translation tools to translate texts in the domain of technology, 31.3% in the domain of travelling (cities, countries), 8.3% for acquiring information on conferences, 20.8% for translating e-mails, 69.5% for translating domain specific texts, and 22.5% for translating texts related to the Internet services, games, wiki articles, literature and for homework translation tasks.

The English-Croatian (En-Cro) translations have been evaluated by 50 native speakers – students of languages, linguistics and information sciences (58% undergraduate, 42% graduate).

Tools and resources

Fig. 1 presents the average grade assigned to Croatian language tools and resources on the Internet and language resources irrespective of the language, i.e. online dictionaries, terminology databases, glossaries, translation memories and translation tools. This grading has been made prior to the pilot research and is based solely on previous experience. The average grade for the Croatian language tools and resources is 3.00, and for language tools and resources in general 3.57.
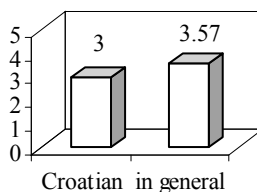

Fig. 1. Average grades for free language resources on the Internet

Fig. 2 presents the average grades for Croatian language tools and resources (online dictionaries, terminology databases, translation memories and GT). The

335

average grade is 2.90, being very close to the general perception of the quality of available tools and resources for Croatian (3.00) given in Fig. 1. The best grade is given to GT (3.54). Fig. 3 presents the average grades assigned to the selected online translation services. The average grade for all four services is 3.325, being close to the average grade 3.57 assigned to free online tools and resources presented in Fig. 1.
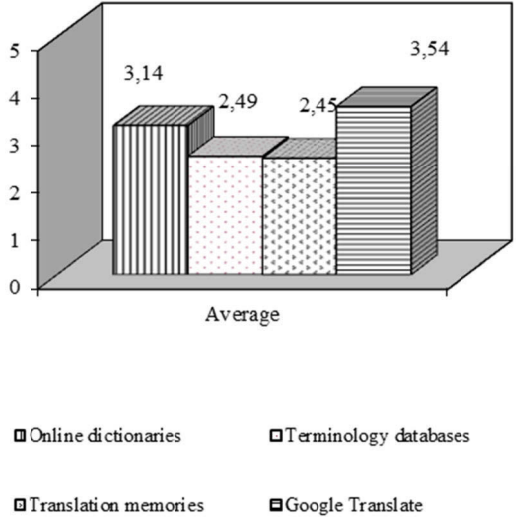

Fig. 2. Average grades for online free Croatian language tools and resources
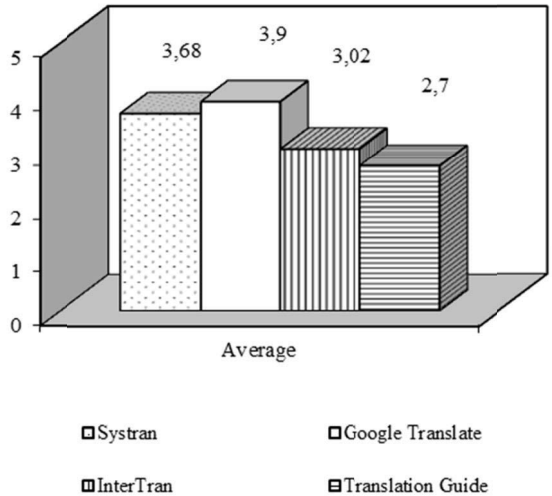

Fig. 3. Average grades for four elected online translation services

Among the interviewed students, 90% of them would like to use the following Croatian tools and resources of the appropriate quality: online dictionaries (90%), MT systems (78%), translation memories (38%), terminology databases (28%), glossaries (24%) and speech-to-text systems (14%) (Fig. 4).
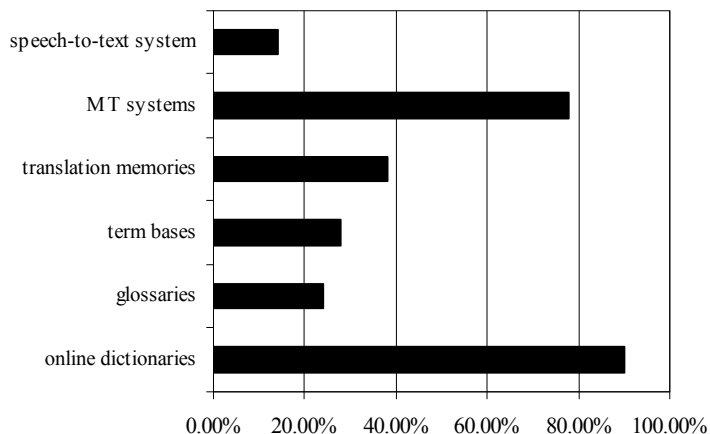

Fig. 4. Desirable resources of the appropriate quality

Results

The evaluation has been made according to the criteria of fluency (indicating how much the translation is fluent in the target language) and adequacy (indicating how much of the information is adequately transmitted). The Cro-En translations have been obtained from four translation services available for the Croatian language (GT, S21, IT, TG), while En-Cro translations have been obtained from GT.

Table 2 presents average grades per each domain in both translation directions (the average of fluency and adequacy).

Table 2. Average grades in Cro-En / En-Cro translations

	Cro-En					Eng -Cro
	GT	**S21**	**IT**	**TG**	**All**	**GT**
city	4.33	4.43	2.18	1.23	3.04	4.71
law	4.63	4.78	2.21	1.13	3.19	4.26
football	4.84	4.72	1.75	1.16	3.12	3.75
monitors	4.72	4.72	1.87	1.15	3.12	4.50
Average	**4.62**	**4.66**	**2.02**	**1.17**	**3.04**	**4.29**

Fig. 5 is a graphical representation of Cro-En translation grades presented in Table 2.

Cro-En translation have been given either low grades (TG and IT) or high grades (S21 and GT), in comparison to the average values. S21 service has obtained slightly better overall average grade than GT (4.66 versus 4.62), better grade in the domain of city description (4.43 versus 4.33), as well as in the legal domain (4.78 versus 4.63). GT has outperformed S21 in the domain of football (4.84 versus 4.72). The average grade for the domain of monitors (4.72) is equal for both services. The services IT and TG have obtained below the average grades in all domains. The best average results for all the services have been obtained in the domain of law (3.19), followed by the domains of monitors and football (3.12), while the lowest grades have been assigned to the city description (3.04), which has the most freedom in style.

Fig. 6 graphically represents grades assigned to En-Cro and Cro-En translations obtained from GT.

En-Cro translations have lower average results than the reverse direction (4.29 versus 4.62), especially in the football domain (3.75 versus 4.84), but also in the domain of law (4.26 versus 4.63) and monitors (4.50 versus 4.72). En-Cro translation direction has higher average grade only in the city description domain which contains shorter sentences, mostly nominative constructions and uses frequent phrase constructions. The most evident difference in the quality of the translation between En-Cro and Cro-En directions is in the text on football, which contains domain specific terms and mostly non-nominative constructions. This causes different types of errors.
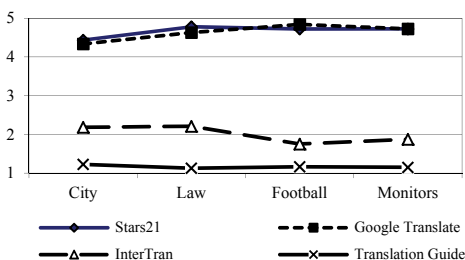

Fig. 5. Average scores for 4 free online translation services for Cro-En
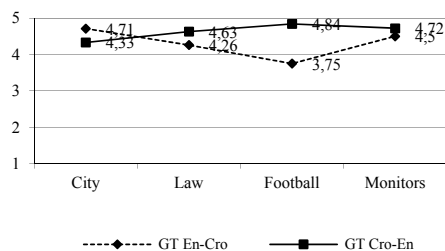

Fig. 6. Average scores for GT in En-Cro and Cro-En directions

Error analysis

The evaluation has been enriched by the analysis of translation errors, which have been divided into several categories:

- morphological errors, i.e. mistakes regarding wrong inflection/endings
- untranslated words
- lexical errors, i.e. wrong word choices, incorrect translations, typing errors, translations of words when not necessary, mistakes that stem from wrong word choices or forms of lexical entries, word omissions

- syntactic errors, i.e. incorrect word order or incorrect punctuation

The following excerpts present Cro-En translations in the domain of football, where there is a significant difference between translation directions, and in the domain of monitors, where the results are similar in both translation directions.

Cro-En
Football

Original: Ovaj Pravilnik utvrđuje globalna i obvezujuća pravila u svezi statusa igrača, njihovo pravo nastupa za sudjelovanje u organiziranom nogometu i njihove transfere između klubova koji pripadaju različitim savezima.

GT: These Regulations establish global and binding rules concerning the status of players eligible to play for their participation in organized football and their transfer between clubs belonging to different federations.

S21: These Regulations establish a global and binding rules concerning the status of players, their right to participate to participate in organized football and their transfer between clubs belonging to different federations.

IT: Ovaj Pravilnik in band their is performing, the law for taking part to organized football which are belonging and their the different coalitions the transfers between clubs.

TG: Ovaj Legality utvrđuje fine plus obvezujuća Pravda into a union fit on playing njihovo pravo intervention for coagency into a organized football plus their transfere između klubova who appurtenant različitim savezima.

Monitors

Original: Pročitajte do kraja ove upute prije nego počnete koristiti ovaj uređaj.

GT: Read these instructions before using this device.

S21: Read these instructions before using this device.

IT: Read our in full ove instructions before počnete are you using this unit.

TG: Pročitajte in full ove instructions before počnete are you using this uređaj.

Translations offered by GT and S21 are very similar, although not identical. GT uses the following phrase ...*players eligible to play for their participation* and S21 the phrase ...*players, their right to participate to participate in*. IT and TG services have not correctly translated isolated words, such as *Regulation*, but have left the source language word *Pravilnik* or have used the incorrect translation *Legality*; the word *federations* has been translated as *coalitions* or left in the source language *savezima*. Instead of the phrase *establish global and binding rules*, the IT has offered *in band their is performing, the law*. TG has produced the translation *utvrđuje fine plus obvezujuća Pravda* with some of the words untranslated (*utvrđuje, obvezujuća*). TG translations are completely incomprehensible and inadequate.

In the domain of monitors, GT and S21 have produced the same output and obtained the highest grades. IT service has produced *before počnete are you using* instead of *before using*. Not only that part of the phrase has not been translated (*počnete*), but there are also syntactic errors. Translation by TG is even worse, because everyday words such as *read, these, device* have not been translated.

Although the translations offered by TG are powered by IT, they differ in the number of untranslated words, i.e. TG does not recognize words with diacritics.

The following excerpts present En-Cro translations in all four domains performed by GT.

City description
Original: It lies on the intersection of important routes between the Adriatic coast and Central Europe.
GT: Nalazi se na sjecištu važnih prometnica između jadranske obale i srednje Europe.
Law
Original: Pursuant to Article 88 of the Constitution of the Republic of Croatia, I hereby issue this DECISION ON THE PROCLAMATION OF THE ELECTRONIC SIGNATURE ACT
GT: Na temelju članka 88. Ustava Republike Hrvatska, donosim ovaj ODLUKA O proglašenju Zakona o elektroničkom potpisu
Football
Original: Registration Period: a period fixed by the relevant Association in accordance with Article 6.
GT: Registracija razdoblje: razdoblje koje utvrdi relevantne Udruge u skladu s člankom 6.
Monitors
Original: Read these instructions completely before using the equipment.
GT: Pročitajte ove upute prije korištenja u potpunosti opreme.

The sentence on the city description has been correctly translated. In the law domain, morphological errors have been made (*na temelju Ustava Republike Hrvatska*), i.e. the genitive form should have been used (*Hrvatske*). In the phrase *donosim ovaj ODLUKA,* the transitive verb should take the direct object in accusative (*donosim ovu ODLUKU*). There is also a mismatch in gender. The phrase *Registration Period* in the text on football might have been translated as a multiword unit (*Registracijsko razdoblje*), in order to avoid a morphological error (*Registracija razdoblje*). Subject-verb gender agreement errors also fall into the category of morphological errors (*razdoblje koje utvrdi relevantne Udruge)* instead of (*razdoblje koje utvrđuje relevantna Udruga* or *razdoblje koje je utvrdila relevantna Udruga*). The verb *fixed* could have been translated in the present (*utvrđuje*) or in the past (*je utvrdila*). In any case, there is either a morphological error or the omission of the auxiliary verb. In the translation of the text on monitors, the translation of *completely* as *u potpunosti* is correct, but stands in the wrong place, which is an example of a syntactic error.
The error analysis in En-Cro translations has shown the highest number of lexical errors, including also errors in style (average 2.44), followed by untranslated words (1.83), morphological (1.75) and syntactic errors (1.38).
The highest number of errors has been found in the text from the football domain (mostly lexical errors and untranslated words). That text has also gained the lowest average grade (3.75).
The best score has gained the translation from the city description domain, although it has the second highest number of all error types. In the city description domain prevail lexical errors.

The lowest number of errors has been found in the legal domain, where all types of errors are evenly distributed.

Morphological errors have been mostly found in the domain of monitors, which has, despite of this, gained the second highest average grade (4.5). The smallest number of morphological errors has been found in the city description domain.

Untranslated words are by far mostly found in the football domain which has gained the lowest average grade.

The dominant value in En-Cro translations, i.e. the number of errors according to the majority, is 1 morphological error in the domains of city description and monitors, while the dominant value is 3 in the legal and football domains. The dominant value with regard to the lexical errors in the city description is 1 and between 1 and 3 in other domains. The dominant value with regard to the number of untranslated words is 1 in all domains. The dominant value with regard to the syntactic errors is 1 in all domains but football, where it is evenly distributed between 1, 2, and 3.

Pearson's correlation

The Pearson's correlation between the number of errors and the average grade is negative, indicating that smaller number of errors augments the average grade, which is mostly reflected in the correlation between untranslated words and average grades.

The correlation between errors types and the criteria of fluency and adequacy has shown that the criterion of fluency is more affected by the increase of lexical and syntactic errors, while adequacy is more affected by untranslated words.

Fleiss' kappa

Fleiss' kappa has been used for assessing the reliability of agreement among raters when giving ratings to the sentences (1). It indicates the extent to which the observed amount of agreement among raters exceeds what would be expected if all the raters made their ratings completely randomly.

$$\overline{P} = \frac{1}{Nn(n-1)} (\sum_{i=1}^{N} \sum_{j=1}^{k} n_{ij}^2 - Nn) \qquad (1)$$

The score is between 0 and 1. A value of 1 implies perfect agreement while values less than 1 imply less than perfect agreement. Interpretation of values used in this case is as follows: < 0 poor agreement, 0.0-0.20 slight agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80 substantial agreement, and 0.81-1.00 almost perfect agreement.

There is a relatively high level of the agreement among raters per domain and per system in Cro-En translations, as given in Table 3. It varies from moderate

(mainly for IT translation service), through substantial agreement (S21 and GT), up to almost perfect agreement (TG).

Table 3. The level of agreement per domain and per system for Cro-En services

Cro-En	city	law	football	monitors
S21	0.61	0.65	0.73	0.70
GT	0.73	0.58	0.61	0.70
TG	0.94	0.99	0.94	0.94
IT	0.52	0.34	0.52	0.51

Table 4. The level of agreement per domain with regard to the criteria of fluency and accuracy for En-Cro translations by GT

EN-CRO	fluency	adequacy
city description	0.58	0.67
monitors	0.53	0.59
law	0.40	0.49
football	0.35	0.37

Table 4 presents the inter-rater agreement for En-Cro translations. The lowest level of agreement has been detected in the domains of football and law, which contain longer and more complex sentences. For the domain of football fair agreement has been detected. The agreement is moderate for the domains of law and monitors and substantial for the city description domain. The level of inter-rater agreement is lower for En-Cro translations in all domains.

Conclusion

This paper presents an evaluation study of machine translations from four domains. Cro-En translations have been obtained from four free online translation services. En-Cro translations have been obtained only from GT. The results regarding the use of freely available language resources indicate that there is a high interest in their use for the Croatian language.

Fleiss kappa shows substantial, even perfect agreement in the evaluation of four translation services. It shows almost perfect agreement in the ranking of TG as the worst translation service. Substantial agreement is achieved for S21 and GT services, which have gained the highest grades. Moderate agreement is shown for IT, which has performed slightly better than TG.

In Cro-En translations the average evaluation results for S21 and GT range from 4.63 to 4.84 for the domains of football, law and monitors. The average grade for the city description in Cro-En translation is lower than in the opposite direction due to more freedom in style.

In En-Cro direction, the translations have been performed by GT and have obtained lower grades than in the opposite direction. This is true for all but the city description domain, which contains mostly nominative constructions, frequent words, and no domain specific terms.

Error analysis shows that the translation grades are mostly influenced by untranslated words (especially the criteria of adequacy), while morphological and syntactic errors reflect grades in smaller proportion.

GT service, which has been used in both translation directions, harvesting data from the Web, seems to be well trained and suitable for the translation of frequent expressions. However, it does not perform well where language information is needed, e.g. gender agreement. The use of background terminology database of multiword expressions and/or translation memory database, would probably improve results, especially translations of specific terms and idiomatic expressions.

Further research in a specific domain would enable a more detailed analysis of specific language phenomena and error types and would enable identifying language-specific problems in automatic MT.

Acknowledgments

References

Agarwal, A.; Lavie, A. METEOR, M-BLEU and M-TER: Evaluation Metrics for High-Correlation with Human Rankings of Machine Translation Output. // *Proceedings of the ACL 2008 Workshop on Statistical Machine Translation*. 2008, pp. 115-118

Aiken, M.; Ghosh, K.; Wee, J.; Vanjani, M. An Evaluation of the Accuracy of Online Translation System. // *Technology, Communication of the IIMA*. 2009.

Aiken, M.; Vanjani, M. B.; Wong, Z. Measuring the Accuracy of Spanish to English Translations. // *Issues in Information Systems*. 7 (2006), 2, pp. 125-128.

Amancio, D. R.; Nunes, M. G. V.; Oliveira Jr., O. N.; Pardo, T. A. S.; Antiqueira, L.; Costa, L. da F. Using Metrics from Complex Networks to Evaluate MT. // *Physica A*. 390 (2011), 1, pp. 131-142.

Bar-Hillel, Y. The Present Status of Automatic Translation of Languages. // *Readings in Machine Translation*. Boston: MIT Press, 2003, pp. 45-77.

Denkowski, M.; Lavie, A. Extending the METEOR Machine Translation Evaluation Metric to the Phrase Level. // *Proceedings of NAACL/HLT*. 2010, pp. 250-253.

Denkowski, M.; Lavie, A. Choosing the Right Evaluation for Machine Translation: an Examination of Annotator and Automatic Metric Performance on Human Judgment Tasks. // *Proceedings of AMTA*. 2010.

Dis Brandt, M. Developing an Icelandic to English Shallow Transfer Machine Translation System, Ms. Thesis. Reykjavik University, 2011.

Doddington, G. Automatic Evaluation of Machine Translation Quality Using Ngram Co-occurrence Statistics. // *Proceedings of the 2nd International Conference on Human Language Technology Research*. 2002, pp. 138–145.

Kit, C.; Wong, T. M. Comparative Evaluation of Online Machine Translation Systems with Legal Texts. // *Law Library Journal*. Vol. 100 (2008), 2, pp. 299–321.

Garcia-Santiago, L.; Olvera-Lobo, M. D. Automatic Web Translators as Part of a Multilingual Question-Answering (QA) System: Translation of Questions. // *Translation Journal*. Vol. 14 (2010), 1.

Giménez, J.; Màrquez, L. Asiya: An Open Toolkit for Automatic Machine Translation (Meta-) Evaluation. // *The Prague Bulletin of Mathematical Linguistics*. 2010, 94, pp. 77–86.

Hampshire, S.; Porta Salvia, C. Translation and the Internet : Evaluating the Quality of Free Online Machine Translators. // *Quaderns: revista de traducció*. 2010, 17, pp. 197-209.

Lavie, A.; Agarwal, A. METEOR: An Automatic Metric for MT Evaluation with High Levels of Correlation with Human Judgments. // *Proceedings of the ACL 2007 Workshop on Statistical Machine Translation*. 2007, pp. 228-231.

Monti, J.; Barreiro, A.; Elia, A.; Marano, F.; Napoli, A. Taking on new challenges in multi-word unit processing for machine translation. // *Proceedings of the 2nd Workshop on Free/Open-Source Rule-Based Machine Translation*. 2011, pp. 11-19.

Papineni, K., Roukos, S., Ward, T., Zhu, W.: BLEU: a Method for Automatic Evaluation of Machine Translation. // *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*. 2002, pp. 311–318.

Somers, H.; Gaspari, F.; Niño, A. Detecting Inappropriate Use of Free Online Machine Translation by Language Students - A Special Case of Plagiarism Detection. // *Proceedings of the 11th Annual Conf. of the European Association for Machine Translation*. 2006, pp. 41-48.

White, J. S.; O'Connell, T. A. Adaptation of the DARPA Machine Translation Evaluation Paradigm to End-to-End Systems. // *Proceedings of AMTA-96*. 1996.

White, J. S.; O'Connell, T.; O'Mara, F. 1995. Evaluation Methodologies in the ARPA Machine Translation Initiative. // *Proceedings of AIPA95*. Automatic Information Processing Association Steering Group, 1995.

Zervaki, T. Online Free Translation Services. // *Proceedings of the 24th Int. Conf. on Translating and the Computer.* London: Aslib, 2002.

A ReSTful Web Service for Multilingual LRT

Kamel Nebhi

LATL - University of Geneva

2, rue de Candole CH-1211 Genève 4, Switzerland

kamel.nebhi@unige.ch

Summary

Today, almost all language tools and resources are available by download only. Someone who wants to use and combine tools and resources are limited. In this context, the establishment of a Web service dedicated to Natural Language Processing seems inevitable. This paper presents a Web service based on a "RestFul" architecture allowing use of the Fips parser, the ITS-2 translation system and the FipsVox text-to-speech system. Users can access all of these technologies by using just a browser while developers can create Web applications based on these resources.

Key words: Language Resources and Tools, Natural Language Processing, Web services, ReSTful Architecture

Introduction and motivation

For years, researchers have been trying to develop language resources and tools (LRT). Almost all of these technologies are often available by download only and/or restricted to a particular platform or environment.

In fact, the "democratization" of these LRT necessarily requires the creation of Web Services (WS).

WS dedicated to Natural Language Processing (NLP) would allow:

- easy access to LRT
- interoperability between different tools and resources
- development of Web application, for example eLearning application (Goldman et al., 2010)

This paper reports our experience integrating different tools and resources into a ReSTful Web Service. Our services provide access to the multinlingual Parser Fips (Nerima and Wehrli, 2009) (Wehrli, 2007) ,the translation system ITS-2 (Wehrli and Nerima, 2008) and the text-to-speech system FipsVox (Goldman et al., 2001).

From now on, this paper is organized as follows: Section 2 describes tools and resources integrated in the WS. In Section 3, we describe the characteristics of the WS and the available LRT. Finally, Section 4 is devoted to the creation of Web applications based on the WS.

Integrated Tools and Resources

This Web Service aims to use several LRT developed at the *Laboratoire d'Analyse et Technologie du Langage* (LATL) of Geneva. Fips (Nerima and Wehrli, 2009) (Wehrli, 2007), ITS-2 (Wehrli and Nerima, 2008) and FipsVox (Goldman et al., 2001) are the main tools of the infrastructure.

The Fips Parser

Fips is a robust multilingual parser which is based on generative grammar concepts for its linguistic component and object-oriented design for its implementation. It uses a bottom up parsing algorithm with parallel treatment of alternatives, as well as heuristics to rank alternatives.

The syntactic structures built by Fips are all of the same pattern, that is: [$_{XP}$ LX R], where L stands for the possibly empty list of left constituents, X for the (possibly empty) head of the phrase and R for the (possibly empty) list of right constituents. The possible values for X are the usual part of speech Adverb, Adjective, Noun, Determiner, Verb, Tense, Preposition, Complementizer, Interjection.

The parser makes use of 3 fundamental mechanisms: projection, merge and move.

Fips builds the canonical form for a sentence, in which extraposed elements (relative pronouns, clitics, interrogative phrases etc.) are coindexed with empty constituents in canonical positions (i.e., typical argument or adjunct positions). For instance, the sentence in (1) below is assigned by Fips the syntactic structure in (2), in which the canonical position of object for the verb address is taken by the empty constituent **e**. The latter stands for the trace of the noun issue, which has been extraposed through relativization. The trace **e**, the relative pronoun φ (a zero-pronoun), and the noun issue are all linked via the index *i*.

1. This too is an issue the Convention must address
2. [$_{TP}$ [$_{DP}$ This] [$_{VP}$ [$_{AdvP}$ too] is [$_{DP}$ an [$_{NP}$ issue$_i$ [$_{CP}$ [$_{DP}$ φ$_i$] [$_{TP}$ [$_{DP}$ the [$_{NP}$ Convention]] must [$_{VP}$ address [$_{DP}$ **e** $_i$]]]]]]]]]

Evaluation

The Fips multilingual parser has been developed for 6 languages (English, French, German, Italian, Spanish and Greek). A comparative evaluation has been conducted to show how the various implementations of Fips compare with respect to a near identical corpus, the European Parliament corpus (Koehn, 2005). We parsed approximately 1 million words in each of the six languages.

The table 1 shows the results:

Table 1 – Comparative evaluation of the parsers

Language	German	English	Spanish	French	Greek	Italian
Number of symbols	1106559	1075246	1228240	1350522	343461	1181785
Unknown words	10685	5852	9165	5643	6788	9006
Number of sentences	47058	41488	46216	45694	13328	44124
% of complete analyses	66,54%	75,68%	71,4%	75,97%	51%	70,76%
Speed (word/second)	17	38	132	83	196	112

ITS-2: the interactive translation system

ITS-2 is a large-scale translation system developed at the LATL of Geneva.

The language pairs currently supported are: French-English, English-French, French-German, German-French, Italian-French, Spanish-French.

The system is based on the familiar transfer architecture, with its three main components, parser, transfer and generation. The parser – associates with an input sentence a set of syntactic structures corresponding to GB S-structures, i.e. surface structures enriched with traces of moved elements and other empty categories. The role of the transfer component is to map source structures onto target structures. Transfer, which occurs at the D-structure level, is to a large extent a matter of lexical correspondence. For each lexical head of a SL structure, the lexical transfer component consults the bilingual lexicon to retrieve the most appropriate TL item, which is then projected according to the X-bar specifications of the TL. Applied recursively over the whole SLD-structure, this process determines an equivalent TL D-structure. From these structures, the generation component derives well-formed S-structures, which are finally converted into the target sentence by morphological process.

At the software level, an object-oriented design has been used, similar to the design adopted for the Fips multilingual parser on which it relies (Wehrli, 2007). To a large extent, ITS-2 can be viewed as an extension of the parser. It relies heavily on the detailed linguistic analysis provided by the parser for the supported languages, and exploits the lexical information of its monolingual lexicons. Both systems aim to set up a generic module which can be further refined to suit the specific needs of, respectively, a particular language or a particular language-pair.

To take a simple example, the direct object of the French verb *regarder* in (3) will be transferred to English as a prepositional phrase headed by the preposition at, as illustrated in (5).

This information comes from the lexical database. More specifically, the French-English bilingual lexicon specifies a correspondence between the French lexeme [$_{VP}$ regarder NP] and the English lexeme [$_{VP}$ look [$_{PP}$ at NP]]. For both sentences, we also illustrate the syntactic structures as built by the parser and/or the generator of ITS-2:

3. Paul regardait la voiture.
4. [$_{TP}$ [$_{DP}$ Paul] regardait$_i$ [$_{VP}$ e$_i$ [$_{DP}$ la [$_{NP}$ voiture]]]]

5. Paul was looking at the car.
6. [$_{TP}$ [$_{DP}$ Paul] was [$_{VP}$ looking [$_{PP}$ at [$_{DP}$ the [$_{NP}$ car]]]]]

Evaluation
The last evaluation was for the Fifth and Sixth Workshop on Statistical Machine Translation. LATL participated in the French-English task in both directions.
The table 2 shows the results in terms of BLEU and translation edit rate (TER) using the newstest2010 and newstest2011 corpus as evaluation set.

Table 2 - Translation results from French to English and English to French measured on newstest2010 and newstest2011

Pair of language	BLEU	TER
French-English	16,5	0,785
English-French	20,4	0,690

FipsVox: a French TTS based on a syntactic parser
FIPSVox is a text-to-speech system for French developed at LATL. It is based on FIPS which produces detailed analyses and the MBROLA diphones-concatenation synthesizer. The syntactic information provided by the parser is directly exploited by the grapheme-to-phoneme module to handle heterophone homographs as well as French elision, denasalisation and liaison phenomena. The prosody generation module also uses this information to determine the dependency between phrases, the accentuation of syllables, and to identify particular syntactic structures such as extraposed constructions (cleft, heavy-NP shift, left dislocation structures, etc.), and parentheticals to derive of appropriate prosodic patterns.

Evaluation
There is no evaluation available for this tool.

Characteristics of the Web Service
A ReSTful Web Service
The services are created on the architectural model ReST. The notion of "Representational State Transfer" (ReST) was introduced in 2000 by Roy Fielding (Fielding, 2000). ReST is an architectural style simply based on the World Wide Web (WWW).
ReSTful means Web Services built using HTTP, URIs, XML, JSON, etc. It's interesting to deploy a WS with a range of existing infrastructure like web server, client library, proxy server, firewall, etc.

ReST architecture is totaly integrated to the WWW that's why it is the simplest and least expensive to implement.

ReSTful approach is basically composed of four concepts:

- the use of the "Uniform Interface" : all resources can be manipulated using the HTTP protocol and the method : PUT, GET, HEAD, POST, DELETE.
- the identification of resources via URI (Uniform Resource Identifier) : each resource can be uniquely identified and adressed.
- the operation are stateless.
- the use of standards like HTML, XML or JSON.

In short, ReST is not a standard it's an architectural style that makes maximum use of web technologies.

Available Resources

Our services provide access to the multilingual Parser Fips, the machine translation ITS-2 and the speech synthesizer FipsVox.

At the moment, most of these tools are available for German, English, Spanish, Greek and French.

Access to resources – Examples

The Data representation is done using standards (UTF-8, XML, TEI, etc.). Applications can be operated directly by URI. Tables 3, 4 and 5 describes resources and request parameters. Picture 1 gives an example of request and response of the *Analyze* resource.

Table 3: Available Resources

Resources	URI	Method	Description
Analyze	baseURI/Analyze[1]	GET	This resource generates a representation of linguistic analysis using the multilingual Parser Fips.
Translate	baseUri/Translation	GET	This resource generates a translation using the machine translation Its.
Speech	baseUri/Speech	GET	This resource generates a speech synthesis of a sentence using FipsVox.

Table 4: Request parameters for *Analyze*

Parameters	Value	Description
ln	fr, en, de, it, es, el (required)	Language for analyze (French, English, German, Italian, Spanish, Greek)
text	String (required)	Text to analyze

---

[1] Currently, baseURI is: http://129.194.19.89

Table 5: Request header for *Analyze*

Header	Value	Description
Accept	Parser, Xml, XmlTei, Tagger	Specify types of content which are acceptable for the response. Values correspond to different representation and format.

```
#Request :
GET /Analyze HTTP/1.1
Host : baseURI
Accept : XmlTei
text = le+chat+mange&ln=fr
#Response :
HTTP/1.1 200 OK
Content-Type : application/xml;charser=UTF-8
<TEI xmlns='http://www.tei-c.org/ns/1.0">
<s xml:lang="french">
    <phr type="DP"  function="SUBJ">
        <w type="DETERMINANT-DEFINI SIN MAS" lemma="le">Le</w>
        <w type="NOM-COMMUN SIN MAS" lemma="chat">chat</w>
    </phr>
    <phr type="" function="Predicate">
        <w type="VERBE-IND-PRE 3 SIN" lemma="manger">mange</w>
    </phr>
</s>
</TEI>
```

Picture 1 - Example of request and response for Analyze

Description of the Web Service

We use the WADL[2] specification to describe our application in a simple format. In this file (cf. Picture 2), we are defining:

- methods and parameters
- responses
- protocols and formats
- URIs of service

---

[2] WADL (Web Application Description Language) is an XML-based file format that provides a machine-readable description of HTTP-based web applications.

```
<resource path="Analyze">
    <method name="GET">
        <request>
          <param name="ln" type="language" style="query" required="true">
            <doc xml:lang="en" title="ln">[language]</doc>
            <option value="fr"/>
            <option value="en"/>
            <option value="de"/>
            <option value="it"/>
            <option value="es"/>
          </param>
          <param name="text" type="xsd:string" style="query" required="true">
            <doc xml:lang="en" title="text">[xsd:string]</doc>
          </param>
          <param name="Accept" type="xsd:string" style="header" required="true">
            <doc xml:lang="en" title="Accept">[xsd:string]</doc>
            <option value="Parser"/>
            <option value="Tagger"/>
            <option value="Xml"/>
            <option value="XmlTei"/>
          </param>
        </request>
    </method>
</resource>
```

Picture 2 – WADL Description of Analyse resource

Web Application based on the Web Service
The user interface

It is possible to use the services with an intuitive interface created as a Web 2.0 application. In this interface, the user can analyze, translate or synthesize a text. The user can view them in several formats. (cf. Picture 3)

Picture 3- Screenshot of the User Interface

FipsColor: an eLearning application based on the WS

FipsColor is a Web application based on the *Analyze* resource (Goldman et al., 2010). The Fips parser analyzes a sentence into a syntactic structure reflecting lexical, grammatical and thematic information. The application adapts the structures in terms of constituents as existent in Fips to a grammatical annotation, giving as well a coloured representation.

This tool is used to highlight a particular aspect of French, such as the lexical ambiguity of certain words. Sentence in Picture 4 illustrates this. In this example, "est" is a verb and a noun. The different colors allow to clearly distinguish the correct grammatical category. FipsColor also underline the syntactic function in order to assimilate agreement or a particular structure (for example relative proposition).

This online interactive application can be used freely by teachers and pupils of primary education.
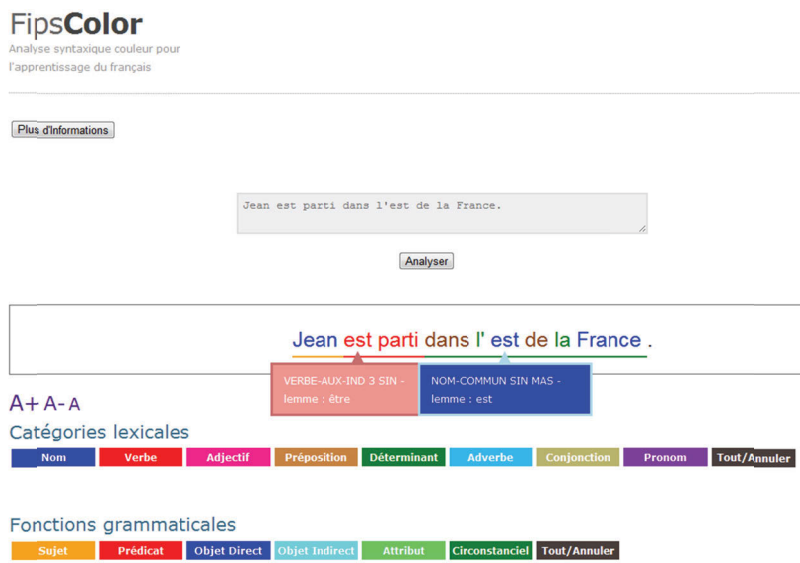
Picture 4 – Screenshot of the Web application FipsColor

Conclusion and Further Work

This paper presented a ReSTful Web Service, which aims at increasing the accessibility and usability of multilingual language resources developed at the LATL of Geneva. We believe that these services can be useful not only for researchers in linguistics, but also for other disciplines which have to analyze corpus. Services also provide creation of Web Application like FipsColor.

In a future version of the Web Service, we would like to include Semantic Web technology in order to provide search and reasoning.

References

Fielding, Roy Thomas. REST : Architectural Styles and the Design of Network-based Software Architectures. Irvine : University of California, 2000

Goldman, Jean-Philippe; Gaudinat, Antoine; Nerima, Luka; Wehrli, Eric. FipsVox : a french tts based on a syntactic parser. // Speech synthesis Workshop / Edinburgh, 2001

Goldman, Jean-Philippe; Laenzlinger, Christopher; Nebhi, Kamel. 2010. Fipscolor : grammaire en couleur interactive pour l'apprentissage du français. // TALN 2010 / Montréal, 2010

Hinrichs, Marie; Zastrow, Thomas; Hinrichs, Erhard. Weblicht: Web-based LRT services in a distributed escience infrastructure. // Seventh conference on International Language Resources and Evaluation (LREC'10) / Valletta, Malta, 2010

Ishida, Toru. Language grid : An infrastructure for intercultural collaboration. // In Proceedings of the IEEE/IPSJ Symposium on Applications and the Internet / Arizona, USA, January. 2006, ,pages 96–100

Koehn, Ph. Europarl: A Parallel Corpus for Statistical Machine Translation. // MT Summit / 2005

Nerima, Luka; Wehrli, Eric. L'analyseur syntaxique Fips. // IWPT 2009 ATALA Workshop : What French parsing systems? / Paris, France, 2009

Schafer, Ulrich. Shallow, deep and hybrid processing with UIMA and heart of gold. // In Proceedings of the LREC-2008 Workshop Towards Enhanced Interoperability for Large HLT Systems : UIMA for NLP / Marrakech, Morocco, 2008, pages 43–50

Villegas, Marta; Bel, Nuria; Bel, Santiago; Rodriguez, Vıctor. A case study on interoperability for language resources and applications // In Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10) / Valletta, Malta, 2010

Wehrli, Eric; Nerima, Luka. Traduction multilingue: le projet multra. // JEP – TALN – RECITAL Conference / Avignon, France, 2008

Wehrli, Eric. Fips, a deep linguistic multilingual parser // ACL 2007 Workshop on deep Linguistic Processing / Prague, Czech Republic, 2007, pages 120–127

Performance Evaluation of Plagiarism Detection Method Based on the Intermediate Language

Vedran Juričić
Department of Information Sciences,
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
vjuricic@ffzg.hr

Tereza Jurić
Department of Information Sciences,
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
tjuric2@ffzg.hr

Marija Tkalec
Department of Information Sciences,
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
mtkalec@ffzg.hr

Summary

This paper presents detection method for source code plagiarism that is based on the intermediate language, and shows its usage in e-learning. Method is tested on the appropriate number of test cases that represent the most frequent code modification techniques. Results and its performance are compared to the existing source code plagiarism detection methods implemented in some of the most known plagiarism detection systems and applications.

Keywords: plagiarism detection, performance, intermediate language

Introduction

Plagiarism is the act of reproducing, or reusing, someone else's work without acknowledging the source [6]. Academic community deals with a huge problem of plagiarism detection, therefore, in order to protect authorship, many algorithms and strategies have been developed. Source code plagiarism detection represents a big problem in educational courses especially in computer science where programming is the main field of work.

Source code is considered plagiarism even if the code does not contain exactly the same elements of someone else's code. Nowadays it's very easy to copy

355

someone else's code from the Internet, whether a function, an algorithm or a complete application. Transformations can be simple, like changing the variable names, modifying comments, and complex, like adding new functions, replacing code structures with equivalents, etc.

This paper deals with the analysis of plagiarism detection systems whose purpose is to detect unoriginal source code in order to maintain copyright infringement.

In this research, three plagiarism detection systems were used. The research was implemented on a group of test cases written in C# programming language. Obtained results were analyzed and compared with the results of algorithm proposed by authors, in order to evaluate the performance of the used systems.

Plagiarism detection method

This paper proposes a method for source code similarity analysis, for .Net programming languages. Method does not analyze the original source code but instead, it analyzes low-level language, called intermediate language.

Intermediate language

All .Net languages are generally compiled twice before finally executed on the operating system. First compiler is language specific, and compiles source code to low-level language called CIL [5] (Common Intermediate Language). For example, code written in C# language is compiled to CIL using C# compiler. CIL is a processor and platform-independent instruction set that can be executed in any environment that supports the Common Language Infrastructure, such as .Net runtime on Windows or cross-platform Mono runtime. CIL code is, upon execution, compiled for the second time using JIT (Just-In-Time) compiler, which generates platform or processor-specific binary code, also known as native code.

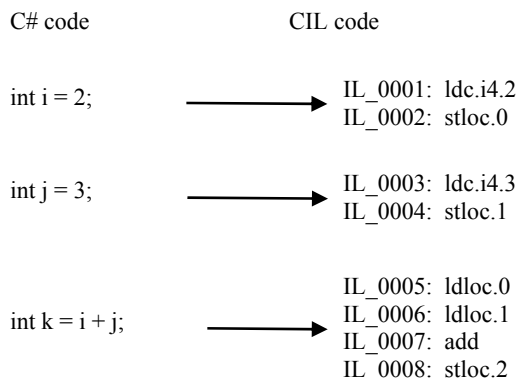| C# code		CIL code
int i = 2;	→	IL_0001: ldc.i4.2 IL_0002: stloc.0
int j = 3;	→	IL_0003: ldc.i4.3 IL_0004: stloc.1
int k = i + j;	→	IL_0005: ldloc.0 IL_0006: ldloc.1 IL_0007: add IL_0008: stloc.2

Figure 1. C# to CIL mapping

Intermediate language code [9] is the lowest-level human-readable programming language, and, therefore, it has lesser commands and simpler structure than third-generation languages like C#. On the other hand, one command of high-level language usually maps to many intermediate language instructions.

Analysis phase

Initial step in similarity analysis is recursive pass through the file system from the specified root folder in order to find all the source files that it contains. It is also possible to specify certain search filters, like allowed extensions and file name pattern.

When initial step completes, all files in the same folder (immediate parent, not the top level) are given to the C# compiler that generates one assembly file per the given folder. Although other methods are possible, application calls the compiler as an external process, and gives specific parameters, like source files and destination file name, via command line interface.

The language of generated assembly is the intermediate language, but its format is not readable and not suitable for analysis, so it must be converted in the text format. This is the third step; a process called disassembling, which is realized in a similar way as previous step, by calling external process and executing ILDasm.exe. ILDasm.exe is a tool that is included in installation of Visual Studio or Framework SDK, which loads an assembly containing intermediate language code and generates a text file.

This text file, generated in the third step, is the actual input file for a comparison. It contains certain lines of code that are not relevant for analysis: metadata and module information, comments (generated by ildasm, not developer), stack related data, etc. Those lines are removed in the fourth, preprocessing step, and not included in the further analysis.

Text file containing disassembler code is parsed line by line and then it is verified if they satisfy predefined patterns. Generally, only lines starting with *IL_* contain intermediate language instructions. The example is shown in Fig. 2 which shows the second instruction (*IL_0002*) inside the observed method, that takes value from the stack and stores it (*stloc*) in the local variable 0 (*.0*).

IL_0002:	stloc	.0
line position	Instruction	variable 0

Figure 2. CIL instruction

The number that follows *IL_* is the position of an instruction inside a method or a property. The exact position of line (instruction) is not relevant for the comparison; what matters is that it exists and occurs, and therefore the whole *IL* part is removed from the line (e.g. *IL_0002*). In the above example, Fig. 2, last num-

ber is the index of local variable. Index is based on the position in the original (high-level language) source code, where the variable is introduced for the first time. This number is also removed from the line, because the order in which appear the variables in the original source code is not relevant for comparison. Also, only variable existence is important. After the fourth step, instruction from Fig. 2 is reduced only to *stloc*.

Comparison phase

When preprocessing phase is completed for all the input assemblies, resulting instruction sets are compared; each processed set is compared to all others sets, and the result is stored in a matrix.

Each line in instruction set can contain one or more elements, but it is taken and compared as one unit, that is, one string. There are numerous methods and algorithms that can be used for string comparison and for calculating similarity between two strings [3], which differ in complexity, calculation time, drawbacks, etc.

The proposed method uses a Greedy String Tilling algorithm [15], which is implemented and used in many of today's plagiarism detection systems. It has worst case complexity $O(n^3)$, but with running Karp-Rabin matching has an experimentally derived average complexity close to linear [15].

Plagiarism detection systems

The following section describes three plagiarism detection systems that were analyzed: MOSS, JPlag and CodeMatch. Their performance was compared with the performance of the algorithm proposed by authors, in order to evaluate the best results obtained by these systems.

MOSS

MOSS (Measure of Software Similarity) is a plagiarism detection software tool developed by Alex Aiken in 1994. MOSS is commonly used in computer science faculties and many other engineering courses. It is provided as a free Internet service hosted by Stanford University and it can be used only if a user creates an account. Files are submitted through the command line and the processing is performed on the Internet server. The current form of a program is available only for UNIX systems.

The program can analyze source code written in 26 programming languages including C, C++, Java, C#, Python, Pascal, Visual Basic, Perl etc. Comparison can be done only between source code files, comparing text files to determine plagiarism between them cannot be done.

MOSS uses Winnowing algorithm based on code-sequence matching and it analyses the syntax or the structure of the observed files.

MOSS maintains a database that stores an internal representation of programs and then looks for similarities between them [10].

The obtained result is displayed in a form of HTML pages or simply in a textual form representing pairs of programs with similar code in an ordered list.

JPlag

JPlag is a free plagiarism detection tool used to detect software plagiarism among multiple sets of source code files. It is commonly used in programming education for detecting unallowed copying of student exercise programs, but it can also be used for detecting stolen software parts among large amounts of source text or modules. JPlag was developed in 1996 by Guido Malpohl and it currently supports C, C++, C#, Java, Scheme and natural language text. Program is available through an installation-free Java Web Start client.

JPlag uses Greedy String Tiling algorithm which produces matches ranked by average and maximum similarity. Average similarity is an average of both program coverages and is the default similarity. If it is big, it indicates that observed programs are working in a very similar way. Maximum similarity is the maximum of both program coverages. It is used to compare programs which have a large variation in size which is probably the result of inserting a dead code into the program to disguise the origin.

Obtained results are displayed as a set of HTML pages in a form of a histogram which presents the statistics for analyzed files.

CodeMatch

CodeMatch is the commercial software included in a CodeSuite collection of analysis tools produced in 2003 by Bob Zeidman and under the licence of a SAFE Corporation. The program is available as a standalone application. It has a free version which allows only one trial comparison where the total of all files being examined doesn't exceed the amount of 1 megabyte of data. The program supports 26 different programming languages including C, C++, C#, Delphi, Flash ActionScript, Java, JavaScript, SQL etc. CodeMatch is mostly used as forensic software in copyright infringement cases exclusively used for source code plagiarism detection.

CodeMatch determines the most highly correlated files placed in multiple directories and subdirectories by comparing their source code. Four types of matching algorithms are used: Statement Matching, Comment Matching, Instruction Sequence Matching and Identifier Matching. These algorithms produce the ranked CodeMatch score which is a combination of all weights given to an each file.

The results come in a form of HTML basic report that lists the most highly correlated pairs of files.

Performance evaluation

Test cases used in this research were written in C# programming language and all of them were created by the authors. Test cases were placed in 6 different

categories and their total number is 50. Categories were constructed considering variable names, types, properties, methods and classes.

All the variable test categories included checking the behaviour of the algorithm when the variable names, types, assigned constant values and location of their declaration varied. Various property definition styles, their type, name and returning values were tested.

Although it tests some of the most common variations with variables, including usage of various existing methods for converting one variable type to another, the syntax itself contains a relatively small number of test cases. Cases that test changing method name, returning type and various parameter reordering, insertions and deletions are also significant parts of the method in question.

The loops category contains test cases that check various loop replacements and definitions while the class category deals with cases that test changing class name, namespace, and reordering and renaming of class members.

Authors conducted manual inspection of all test cases used in the research. The results of the aforementioned manual comparison are shown in a 50x50 matrix the rows and columns of which are correspondent to the test cases used. Each cell represents the value between two test cases. This matrix represents a reference matrix to which all results obtained by plagiarism detection systems are compared.

The two used evaluation methods were precision and recall including their harmonic mean, the F measure. Those methods evaluate the algorithms' behaviour and sensitivity to various code modification techniques. Precision is defined as a fraction of correctly categorized test cases divided by the number of test cases claimed to be similar [12]. Recall is defined as fraction of correctly categorized test cases divided by the number of test cases manually categorized as similar [12]. F measure is defined as a harmonic mean of precision and recall, so that both measures are equally represented [12].

Results

The reference matrix contains only values one and zero, where one indicates that the similarity between test cases is relevant and that they should be treated as similar or equal. Similarity matrices obtained by the plagiarism detection systems contain decimal values in range from zero to one, so they were converted to the suitable values in order to analyse performance. Conversion is based on the threshold, so that all values above threshold are converted to one; otherwise, they are converted to zero.

Authors tested the relation between a threshold value and the calculated precision, recall and F measure at this threshold, which enabled the authors to identify the best values and performance for each plagiarism detection system. Results for each system are presented in the graphs below. Graphs display precision (p), recall (r) and F measure (F) in relation to threshold which is displayed on the horizontal axis.
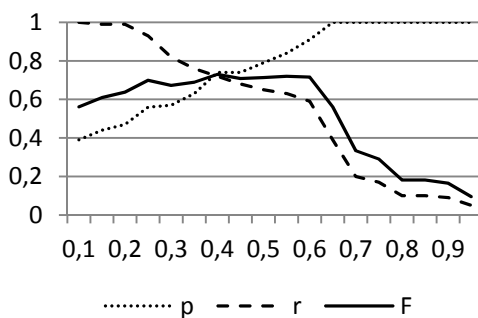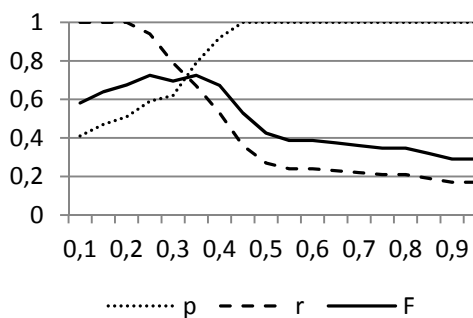
360

Figure 3 - MOSS results
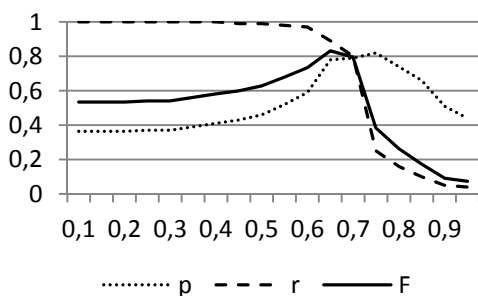

Figure 4 - JPlag results
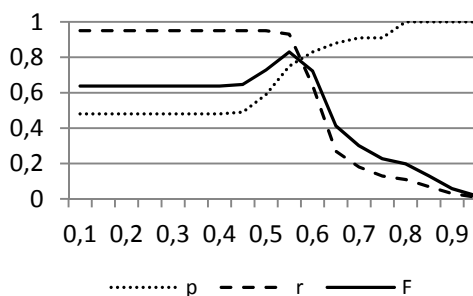

Figure 5 - CodeMatch results

Figure 6 - ILMatch results

As it is expected, all plagiarism detection systems have very high recall when the threshold is low, and it decreases as the threshold increases. ILMatch is the only system that does not reach recall of 100%; its highest value is 95%. On the other hand, precision rises with a threshold, and reaches 100% on the high threshold, except for the CodeMatch system, whose maximal precision is 81%.

The best identified F-measures for tested plagiarism detection systems are shown in the Fig. 7. Performance of MOSS and JPlag is almost the same: the best F-measure for those systems is about 73%, while CodeMatch and ILMatch show the best performance, their best F-measure is about 85%.

By analyzing ILMatch behaviour on individual test cases, the authors concluded that changes in comments do not have impact on similarity, because user comments do not appear in intermediate language code. Also, because source code is not analyzed, modifications of code formatting have no impact on similarity. Modifications to intermediate language code that are made in preprocessing phase, ensure that these transformations do not affect the results of comparison: modifications of variable and class names, changing names of class members, changing data type of variables and constants and changing values of constants.

Replacing expressions and loops with equivalents and changing the structure of selection statements has slight impact on comparison results. Rewriting code in different programming language also has little impact similarity. Transformations that can cause significant differences in calculating similarity are reordering operands in expressions, changing the order of class members, changing the order of statements and adding redundant statements and variables.
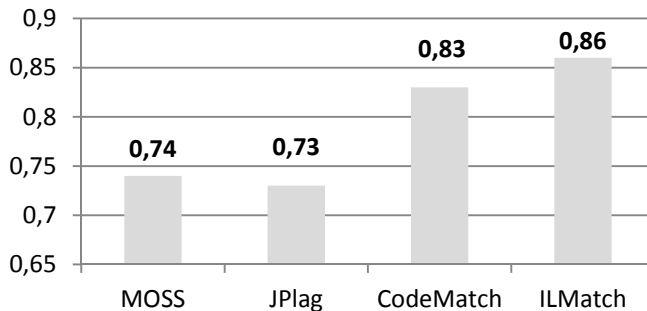

Figure 7 - Highest F-measures

Conclusion

This paper presented a plagiarism detection method based on the intermediate language. The system based on the proposed method was compared with the most used plagiarism detection systems that were available to the authors and that supported the language the test cases were written with: MOSS, JPlag and CodeMatch.

Test cases were designed and written so that they analyse systems' behaviour under different types of code modifications, which are commonly used to mask code reuse and copying. By analysing their performance, authors determined that the system based on the intermediate language showed the best results, that is, it had the best F-measure.

References

[1]   Bowyer K.W, Hall L.O. Experience Using "MOSS" to Detect Cheating On Programming Assignments. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.73.7062&rep=rep1&type=pdf
[2]   Chapman S. SimMetrics, Open source library of Similarity Metrics. 2006. http://staffwww.dcs.shef.ac.uk/people/S.Chapman/stringmetrics.html
[3]   Clough P, Plagiarism in natural and programming languages: an overview of current tools and technologies. 2000. ftp://www.dlsi.ua.es/people/armando/maria/Plagiarism.rtf
[4]   Cohen W.W, Ravikumar P, Fienberg S.E. A Comparison of String Distance Metrics for Name-Matching Tasks. http://www.cs.cmu.edu/~wcohen/postscript/ijcai-ws-2003.pdf

[5]   Common Intermediate Language. 17.12.2010. http://en.wikipedia.org/wiki/
      Common_Intermediate_Language
[6]   Goel S, Rao D et al. Plagiarism and its Detection in Programming Languages.
      http://www.stanford.edu/~drao/Resources/Plagiarism.pdf
[7]   JPlag Detecting Software Plagiarism. 16.6.2010. https://www.ipd.uni-karlsruhe.de/jplag/
[8]   Kwan R, Fox R, Chan F.T. Enchancing learning through technology: research on emerging
      technologies and pedagogies. World Scientific Publishing Co. Pte. Ltd. Singapore, 2008.
[9]   List of CIL Instructions. 17.12.2010. http://en.wikipedia.org/wiki/List_of_CIL_instructions
[10] Plagiarism Detection. 20.6.2011. http://theory.stanford.edu/~aiken/moss/
[11] Prechelt L, Malpohl G, Philippsen M. Finding Plagiarisms among a Set of Programs with
      JPlag. Journal of Universal Computer Science; 2002. http://pswt.informatik.uni-
      erlangen.de/EN/ publication/download/jplag.pdf
[12] Precision and recall. 15.6.2011. http://en.wikipedia.org/wiki/Precision_and_recall
[13] S.A.F.E. Software Analysis & Forensic Engineering Corporation. CodeSuite Version 4.2
      User's Guide. http://www.safe-corp.biz/documents/CodeSuite_Users_Guide.pdf
[14] Schleimer S, Wilkerson D.S, Aiken, A. Winnowing: Local Algorithms for Document
      Fingerprinting; 2003. theory.stanford.edu/~aiken/publications/papers/sigmod03.pdf
[15] Wise M.J. String Similarity via Greedy String Tiling and Running Karp−Rabin Matching.
      1993. http://vernix.org/marcel/share/RKR_GST.ps

Language Identification of Web Data for Building Linguistic Corpora

Marija Stupar
Department of Information Sciences,
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
mstupar@ffzg.hr

Tereza Jurić
Department of Information Sciences,
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
tjuric2@ffzg.hr

Nikola Ljubešić
Department of Information Sciences,
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
nljubesi@ffzg.hr

Summary

In this paper we inspect a series of methods for language identification on web data. We start from the standard two methods based on function word frequencies and Markov chains. We investigate the problem on both the document and the paragraph level. After obtaining an insight in the strengths and weaknesses of these basic methods, we propose two hybrid methods where the more complex one outperforms or performs equally well as the best basic method. Identifying language on each paragraph of more than three million documents collected for the Croatian Web Corpus hrWaC shows that around 96% of the documents are monolingual and that the language distribution, as expected, follows a power-law distribution.

Key words: language identification, Web data, Croatian Web corpus, Markov model, function words

Introduction

The Web represents a freely available and rich source of linguistic material. With a disparate nature of contained sources, it can be used to conduct various types of linguistic research. There is a high possibility of finding texts in more

than one natural language within such sources. The problem of multilinguality therefore presents a challenge.

Language identification is a process which aims to label textual documents according to the language they are written in, and it is often applied to many fields of natural language processing since multilinguality is nowadays a frequent phenomenon.

The goal of this paper is to define a method for identifying the language which the documents collected from the Web are written in. The described method is primarily developed for building hrWac, the Croatian Web corpus, although it can be applied to other problems such as finding texts in different languages on the Web that are translations of each other (building thereby automatically parallel corpora).

In this paper we compare the two main approaches for language identification - the linguistic and the statistical one. The linguistic approach is based on function word distributions while the statistical approach is based on second-order Markov models trained on small language samples of all anticipated languages. After obtaining an insight in the weaknesses and strengths of each approach, we propose two hybrid approaches combining these two methods.

Related work

Using Web resources can be a useful basis for constructing corpora in fields of linguistics, language technologies and translation. Projects such as the WaCky initiative (Baroni et al, 2009.) aim to provide a set of tools to process, index and search the data gathered from the Web.

Language identification is a widely studied field, and many different approaches have been introduced to solve this problem. The methods vary between using special characters, information about short words, frequency of n-grams, Markov models, etc. There are also approaches that combine various methods in order to achieve better results.

The basic approaches used in this research are described in following papers: the linguistic approach based on function words has been studied by (Ingle, 1976) and (Kulikowski, 1991), and some of the Markov model approaches have been presented in (Schmitt, 1991) and (Dunning, 1994).

Experimental setup

The twelve languages observed in the research are Czech (cs), German (de), English (en), Spanish (es), French (fr), Croatian (hr), Hungarian (hu), Italian (it), Polish (pl), Slovak (sk), Slovenian (sl), and Swedish (sv). They were chosen upon their incidence in hrWac obtained through a corpus concordancer.

To have a sense of how hard our problem will be we first studied the similarity of the languages chosen for our experiment. Therefore we used the data found in „The Language Table" by (Crúbadán, 2007). The table shows the cosine similarity between the 3-gram profile vectors for each language. The data for

366

our 12 languages is given in Table 1. We expect to find it harder to distinguish between similar languages, such as Czech and Slovak, Croatian and Slovenian, or Spanish and French. Hungarian, on the other hand, seems very different from other languages, and therefore has a high possibility of being correctly identified. It should be noted that we did not take Serbian or Bosnian into consideration in this research for two reasons:

1. It is not too likely to find significant amount of such material on the Croatian Internet domain
2. Distinguishing these languages should be regarded a separate problem as described in (Ljubešić, Boras, Mikelić, 2007) which should follow the first language identification phase we investigate here

Table 1: A snippet from "The language table"

	cs	de	en	es	fr	hr	hu	it	pl	sk	sl	sv
cs	-	18	22	26	22	53	25	31	42	70	54	23
de	18	-	34	34	35	12	17	31	20	17	18	53
en	22	34	-	27	33	16	16	35	15	17	19	35
es	26	34	27	-	62	22	18	56	18	23	28	38
fr	22	35	33	62	-	18	15	48	15	18	22	35
hr	53	12	16	22	18	-	11	31	39	51	74	24
hu	25	17	16	18	15	11	-	14	10	22	13	21
it	31	31	35	56	48	31	14	-	22	28	38	32
pl	42	20	15	18	15	39	10	22	-	50	40	18
sk	70	17	17	23	18	51	22	28	50	-	55	22
sl	54	18	19	28	22	74	13	38	40	55	-	26
sv	23	53	35	38	35	24	21	32	18	22	26	-

We distinguish two phases in our experiment: 1. identifying language on the document level and 2. identifying language on the paragraph level. In each phase of the experiment we evaluated both approaches – the linguistic and the statistical one. After obtaining an insight into the strengths and weaknesses of every approach on both levels, we propose two hybrid approaches and evaluate these on both the document and paragraph level.

The linguistic approach uses lists of function words from all languages in question and picks that language for which the highest percentage of words could be identified as function words of the respective language.

The statistical approach uses second-order Markov models, i.e. conditional probabilities of a character regarding the two previous characters for which distributions of bigram and trigram characters is calculated on a training set. A detailed overview of the method used is given in (Ljubešić, Boras, Mikelić, 2007).

The data necessary for building both methods was collected by hand. The number of collected function words, i.e. the amount of training data for building Markov models is given in Table 2.

For evaluation purposes we built two gold standards from documents collected for purposes of building hrWaC, one gold standard for each level.

For the document level we collected 20 documents per language. The documents were also obtained with help of a concordancer. We are aware of the fact that this uniform distribution does not follow the actual language distribution on the Croatian web. Since it would be very hard, if not impossible to build a labeled sample of random documents with enough examples for all 12 languages, we were forced to ignore the real distribution and approximate the uniform one. Since a significant amount of documents on the Web is written in more than one language, we included in the sample also documents written in more than one language. To keep the complexity of the task under control, our rule of thumb was to label a document with a specific language label if at least 70% of the document was written in that language. We consider the documents containing less than 70% of any language unsolvable on the document level.

Table 2: Amount of data collected for each basic method (the number of function words per language, and character count as training data for building the Markov model)

Language	Function words	Character count
Czech	210	150601
German	334	150156
English	230	150041
Spanish	217	150926
French	260	150083
Croatian	204	157366
Hungarian	223	152202
Italian	219	150459
Polish	268	150198
Slovak	168	150046
Slovenian	256	143841
Swedish	256	150762

For evaluating the methods on paragraph level we labeled paragraphs in 50 documents by language they are written in. Thereby we labeled 750 paragraphs in total.

Our evaluation measure is accuracy (a+d/a+b+c+d), where the nominator contains the number of correct decisions and the denominator the overall number of decisions made.

Results and discussion

The results of the evaluation of the two standard methods are given in Table 3. All the results are rather high, but it is obvious that Markov model consistently achieves better results. Markov model was identically accurate on both document and paragraph level, while the method using function words achieved

better results on the paragraph level which could be considered rather strange. It is our opinion that this is because of different languages present in a number of documents and the inability of the method to deal with mixed content. On the paragraph level this was not an issue since most of the paragraphs are written in one language only.

A manual evaluation of the results showed the strengths and weaknesses of every method. Markov models are prone to making wrong decisions if a segment in the string contains characters characteristic for another language (a document written in English was recognized as Croatian due to frequent occurrence of the named entity "Sveučilište u Zadru - Odjel za njemački jezik i književnost"). On the other hand, the function words method tends to make wrong decisions in case of an overlap in function words between more languages (a document written in Hungarian was recognized as English, due to occurrence of the same function words with different usage such as "a" in English meaning "the" in Hungarian, or "is" in English meaning "also" in Hungarian) and in case of shorter texts. The function words method, as shown in the automatic evaluation, is in general more prone to errors.

Table 3: Results of the evaluation of the traditional methods

	Function words	Markov model	Function words	Markov model
	Document level		Paragraph level	
Positive	234	239	745	747
Negative	6	1	5	3
Accuracy	0.975	0.996	0.993	0.996

Finally, we propose a hybrid approach that combines the two methods evaluated above having in mind that these methods are erroneous in different situations. The first method calculates the harmonic mean of the certainty of the function words method and the Markov model method (certainty is calculated as $a/(a+b)$ where a is the first result, and b the second best result). The more sophisticated hybrid method takes into account the strengths of each method and thereby does the following:

- If the Markov model and function words method give the same results, the result is accepted
- In case the results of both models are not the same, but the second best result of the Markov model method is identical to the first result of the function words method and its certainty is over 0.6, the result of the function word method is accepted
- Otherwise the result of the Markov model method is accepted

Thereby we change the decision made by the Markov model method only in case the second-best guess of the Markov model method and the first guess of the function word method are identical with a safety margin of 0.1.

369

The results of the automatic evaluation of the hybrid approaches are given in Table 4.

The results show a small, but consistent improvement. What is more interesting, these hybrid methods obviously handle significantly better the case where a mixture of languages is present in the string. On the paragraph level there is no visible improvement when comparing the results to the results obtained by the Markov model method. The question that arises here is if a difference would become visible on a larger (more representative) sample.

Table 4: Results of the evaluation of hybrid methods

	Harmonic balance	Sophisticated method	Harmonic balance	Sophisticated method
	Document level		Paragraph level	
Positive	239	240	746	747
Negative	1	0	4	3
Accuracy	0.996	1.0	0.995	0.996

Because of the improvements shown in specific situations shown by the sophisticated hybrid method, we decided to use this method for the task of identifying languages in our emerging web corpus. We analyzed 3,924,189 documents by each paragraph. In 95.9% of the documents all paragraphs were identified as being written in same language. From all documents containing a paragraph identified as Croatian, 95.8% of the documents were pure Croatian. In Table 5 we give a distribution of languages as identified on paragraph level. The data show a power-lawish distribution where 90% of the paragraphs are written in Croatian. Second-best, as expected, is English with 8%, Slovene with 1% and the remainder of languages making only 1% of the paragraphs.

Table 5: Distribution of languages as identified on paragraph level

Language	Number of paragraphs	Paragraph percentage
Croatian	25347696	89.9%
English	2195590	7.8%
Slovene	288829	1%
German	111078	0.4%
Italian	64268	0.2%
Spanish	39515	0.1%
Swedish	34388	0.1%
French	33817	0.1%
Czech	31812	0.1%
Slovak	22313	0.1%
Polish	18791	0.1%
Hungarian	15404	0.1%

Conclusion

In this paper we have compared the two mostly used language identification methods on web data – the function words method and the Markov model method. We have shown that in general Markov model outperforms the function words method. A case where Markov model fails is if a sequence of characters specific for another language, like a named entity, is found in the data. On the other hand, the function words method underperforms on shorter texts and suffers from collisions of function words between languages. These methods perform very well on paragraph level as well, even outranking the document level results on some occasions since web documents tend to contain mixed language content.

We proposed two hybrid approaches that showed to be more efficient on the document level, i.e. on data containing mixed language content. It is our belief that on a larger gold standard the hybrid methods would outperform the standard methods on paragraph level as well.

In the end we identified the language on paragraph level in documents collected for the Croatian Web Corpus hrWaC and showed that around 96% of documents are written in only one language where the remaining 4% have mixed content. Additionally, we showed that the distribution of languages is power-lawish where Croatian, English and Slovene make 99% of the data.

References

Martins, B; Silva, M.J. Language identification in web pages. // The 20th ACM SAC Symposium on Applied Computing. Document Engeneering Track. / L. M. Liebrock (ed.). 2005, 773 - 777.

Dunning T. Statistical identification of language. // Technical Report MCCS. New Mexico: New Mexico State University, 1994.

Ingle N. A language identification table. The Incorporated Linguist, 15(1976), 4.

Kulikowski S. Using short words: a language identification algorithm. // Unpublished technical report, 1991.

Baroni, M. et.al. The WaCky Wide Web : A Collection of Very Large Linguistically Processed Web-Crawled Corpora. // Language Resources and Evaluation 43(2009), 3, 209-226.

Ljubesic, N; Mikelic, N; Boras, D. Language identification : How to distinguish similar languages? // Proceedings of the 29th International Conference on Information Technology Interfaces. 2007, 541-546.

Vojtek, P; Bielikova, M. Comparing Natural Language Identification Methods based on Markov Processes. // Slovko - International Seminar on Computer Treatment of Slavic and East European Languages. 2007.

Schmitt J.C. Trigram-based method of language identification. US Patent 5062143. 1991.

N-gram Overlap in Automatic Detection of Document Derivation[*]

Siniša Bosanac
Department of Information Sciences,
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
sbosanac@ffzg.hr

Vanja Štefanec
Department of Information Sciences,
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
vstefane@ffzg.hr

Summary

Establishing authenticity and independence of documents in relation to others is not a new problem, but in the era of hyper production of e-text it certainly gained even more importance. There is an increased need for automatic methods for determining originality of documents in a digital environment.

The method of n-gram overlap is only one of several methods proposed by the literature and is used in a variety of systems for automatic identification of text reuse. Although the aforementioned method is quite trivial, determining the length of n-grams that would be a good indicator of text reuse is a somewhat complex issue. We assume that the optimal length of n-grams is not the same for all languages but that it depends on the particular language properties such as morphological typology, syntactic features, etc. The aim of this study is to find the optimal length of n-grams to be used for determining document derivation in Croatian language.

Among the potential areas of implementation of the results of this study, we could point out automatic detection of plagiarism in academic and student papers, citation analysis, information flow tracking and event detection in on-line texts.

Key words: Document derivation, text reuse, n-gram overlap, automatic plagiarism detection, string metrics

---

[*] The preliminary results of this research were in somewhat shorter form presented at the student linguistic conference StuLiKon, held on 6th – 8th May, 2011 in Belgrade, Serbia.

Introduction

The problems of originality and authenticity were always contemporary, but nowadays they demand even more attention. There are two principal factors that have made these problems so important: development of information and communication technology and commodification of intellectual products. Our objective was to deal with only a small segment of this problem that can be examined empirically – derivation of content. In this paper we will establish the basic theoretical framework of document derivation detection and describe our effort to test and modify an existing method for use on natural language texts in Croatian. Text reuse is a very common occurrence in the process of text production and some say that it is as old as storytelling itself.[1] In most cases the motivation to examine it came out of frustration caused by witnessing plagiarism and similar negative phenomena, and the same is valid for this research.[2]

Increased usage of ICT for producing texts and their dissemination has intensified this phenomenon, and also removed some of the technical limitations. Ethical and legal factors are also to be considered when analyzing the problem, but they will not be dealt with in this paper. An example supporting the claim that plagiarizing has become socially acceptable and moreover, profitable, is the existence of on-line markets of essays.[3] Automatic methods for detection of derivation have been researched to a far greater extent for use in the IT sector, where they are used to detect plagiarized programming source code, than for natural language texts.[4] Efforts made for detecting reuse in natural language text were aimed at protecting the content distributed by news-wire agencies, and for designing tools for detection and discouraging plagiarism in academia.[5]

Theoretical framework

Text reuse is a process by which literal content from a single source document is reused in the creation of a target document. Content is reused in the same context either word-for-word (verbatim) or paraphrased (rewritten).[6]

Derivation is the relationship between the two documents in which can be shown, with confidence, that the target document used the source document in its creation.[7]

---

[1] Wilks (2000), according to Clough (2001, p.3).

[2] The authors of this paper have independently dealt with plagiarism in their previous works: Bosanac *et al.* (2009), Štefanec (2010), Bosanac & Štefanec (2011).

[3] For examples refer to: http://seminarski.blog.hr/, http://www.maturskiradovi.net/.

[4] Clough (2001, p.12).

[5] Measuring Text Reuse project at the University of Sheffield was dealing with text reuse in news articles; Turnitin and Plagiarism Detect are tools popular in academia.

[6] Clough (2001, p.27).

We have listed, according to our experience, the most common examples of content derivation. We do not consider this list to be exhaustive.

Examples of content derivation generally considered desirable: 1) *Quoting* and using as sources texts from the domain of scientific and technical literature, literary fiction and journalism; 2) *Document updating*, i.e. adding new content to existing documents, but it can also refer to on-line journalist articles; 3) *Relaying the content* of press releases and content supplied by news agencies in media publications; 4) Automatic and manual *summarization*.

Examples of content derivation generally considered undesirable: 1) *Plagiarism* in all domains, but specifically in academia, journalism and literary fiction; 2) *Non-critical relaying* of content in media, e.g. disseminating information without accrediting the source; 3) *Journalistic theft* which occurs when a journalist copies the content from another media that has legally obtained it through subscription to a news agency's feed.

In our research, we have recognized the following methods of reusing text in a new document: *copy-pasting* (verbatim reuse) and paraphrasing (rewriting). Text considered to be reused verbatim has exactly the same word forms and structure as in the source document, shares the same context in both documents and can be aligned with the corresponding text in the source document.

Text that is reused by rewriting shares some similar word forms between the source and target documents, not all the words are exactly the same due to editing transformations, the context is the same and the rewritten text can be aligned with corresponding text from the source document.[8] For example this is a result that shows text reuse by paraphrasing:

Source document: "***U planinskom području Perua****, daleko od glavnog grada Lime i daleko od bilo kojeg grada*, postoji **mreža seoskih knjižnica**."

Target document: "*U svom govoru Kay Raseroka spomenula je interesantan primjer razvoja ruralnih knjižnica* **u planinskom području Perua*. **Daleko od glavnog grada Lime i daleko od bilo kojeg** *drugog* **grada** *razvila se* **mreža seoskih knjižnica**."[9]

---

[7] Clough (2001, p.28).

[8] Clough (2001, p.29).

[9] Source document: "In the mountainous area of Peru, far from the capital city Lima, and far from any city, there is a network of village libraries."
Target document: "In her speech, Kay Raseroka mentioned an interesting example of development of rural libraries in the mountainous area of Peru. Far from the capital city Lima, and far from any other city, a network of village libraries has developed."

Detection of document derivation
N-gram overlap

In linguistics, the term *n-gram* denotes a sequence of *n* successive language units. Depending on the application, those language units can be letters, syllables, morphemes, words, etc. In this case we will be dealing with n-grams of words. Using the n-gram overlap method we will try to determine the amount of n-grams common to two documents. This method is based on the assumption that overlapping of longer n-grams could be an indicator of text reuse. However, the most representative n-gram length for detecting derivation in a certain language should be determined experimentally because it depends on the morphological and syntactical features of the language. It is necessary to bear in mind that texts in the same language share a certain amount of lexis, and that amount increases substantially if the texts are on the same topic or belong to the same functional style. Therefore, if too short n-grams are taken into account, matching will be higher and nonrepresentational, while if too long n-grams are selected there is a risk that not all cases of derivation will be recognized. Some of the authors dealing with this problem in English are Lyon *et al*. (2001)[10] who showed that word trigrams could be used for discriminating independent from copied text within a collection of texts, Bloomfield[11] who claimed that 6-grams are more representative for identifying collusion between student assignments, Shivakumar and Garcia-Molina (1995) who find unigrams to be optimal for copy detection, etc. From such diverse results it is evident that the problem is far from simple and that we cannot even predict the interval in which our results could fall into. Besides finding the most representative n-gram length, we will try to determine the amount of common n-grams which can be used as a threshold for automatic derivation detection.

N-gram overlap is not the only method that can be used for detection of text reuse, but it is certainly one of the simplest. It lacks precision since it takes only fixed sized n-grams and performs simple binary comparison, but its low complexity makes it suitable to be applied over large collections of documents. In that way, this method can be used for identifying candidates that should be analyzed more thoroughly using some of the more precise and complex algorithms. For example, the widely used Longest Common Subsequence, or somewhat less familiar Greedy String Tiling algorithm. While the first is the basis of various file comparing functions and programs, the latter is specifically designed for detection of plagiarism in computer programs and other texts.[12]

---

[10] According to Clough (2003, p.117).

[11] *Ibid*., p.118.

[12] Wise (1993).

In this research we will perform measurements on a collection of texts in order to determine the relevant parameters for using n-gram overlap method in identifying derivation of documents in Croatian.

Measure

In order to be able to quantify the amount of text reuse in derived documents, it is necessary to introduce measure capable of expressing that kind of relation. Measure commonly used for this purpose is *resemblance*.

Resemblance is a symmetric measure which expresses the amount of common content within the total content of two documents. Resemblance will be calculated using the Jaccard similarity coefficient or Jaccard index. Since Jaccard index is a statistic for comparing the similarity of sample sets[13], our documents will be treated as sets of n-grams. That means that, due to the fact that a set consists only of unique elements, we will be dealing with types of n-grams instead of tokens, i.e. only one occurrence of n-gram will be recorded in the set.[14] Resemblance will then be computed as the size of the intersection divided by the size of the union of n-grams.

$$r(A,B) = \frac{|F(A) \cap F(B)|}{|F(A) \cup F(B)|} \quad [\cdot 100\%]$$

Although it is possible to use weighted resemblance instead, in which case higher weight is given to more frequent n-grams, it will not be used here.

Text collection

The text collection on which we conducted our measurement consisted in total of 238 documents of various sizes (69 – 34,397 words). They were taken from the digital repository of the Library of the Faculty of Humanities and Social Sciences, Web news sites, and other Web sources. Documents in the collection were classified by topic and functional style. The general topic of documents was determined according to classification used at the source, title, keywords, and additionally, for shorter texts, manually. The functional style was determined primarily according to the classification at the source. The collection consisted of 39 diploma papers, 42 scientific articles, 61 news articles, 61 literary columns, and 35 documents classified as "other" from the fields of library science and psychology.

Methodology

For the purpose of this research, we have composed a module in a dynamic programming language using which we have performed measurements over our

---

[13] Wikipedia contributors (2011)

[14] Clough (2001, pp.51-52)

text collection. The module was designed to combine all documents from the text collection into binomial combinations (they will be referred to as *derivation pairs*) and perform identification of common n-grams within them. In that way, out of 238 documents our module produced 28,203 derivation pairs which will serve as our training data set.

$$C_2^{238} = \binom{238}{2} = \frac{238!}{2!(238-2)!} = 28203$$

Documents in every derivation pair were compared in 10 subsequent iterations. In each iteration the comparison was performed with respect to n-gram of different length, starting on the level of 10-grams to the unigram level. In addition to calculating the resemblance in each iteration, the module generated an exhaustive summary containing the list of common n-grams for every derivation pair. After completing the measurements, every derivation pair was given a desired output value, namely, derivation pairs were marked either as derived or non-derived using a semi-automatic method. Specifically, derivation pairs consisted of documents for which there is a negligible likelihood of derivation were automatically marked as non-derived, while those for which non-derivation cannot be assumed, were classified manually by examining the generated summary. In overall, out of 28,203 pairs, derivation was established in 265 of them, while other 27,938 pairs were marked as non-derived. After that, all derivation pairs were ranked according to resemblance respectively for each n-gram length.

Further analysis of the results included finding the resemblance value that yields the maximum F1-measure for each respective n-gram length. Resemblance value will be used as a threshold for automatic classification of derivation pairs, and precision and recall scores will be obtained by comparing the output of the algorithm with the desired output value. F1-measure, as a harmonic mean of precision and recall, will show us for which resemblance value we have the optimal ratio between false-negatives and false-positives. Figure 1 gives an illustration of how F1-measure changes with the resemblance on the 3-gram level.
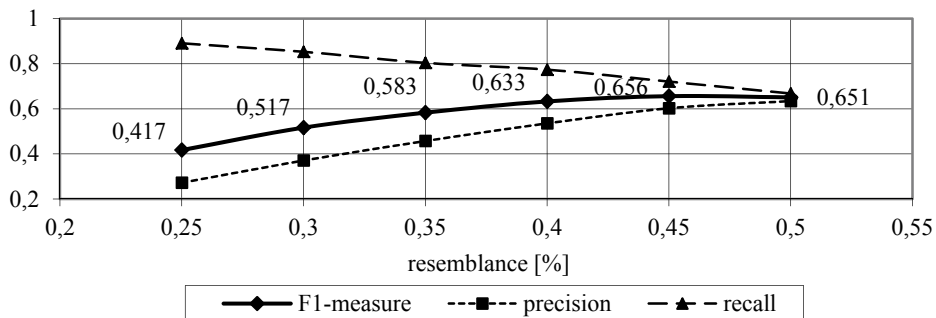

Figure 1: Trend of precision, recall and F1-measure on the level of 3-gram

After finding the F1-measure maxima at the level of each n-gram respectively, by comparing their values, it will be possible to determine at the level of which n-gram F1-measure reaches the highest value. That n-gram will then be considered as the most representative for detection of derivation, and the resemblance which yields that F1-measure value will be considered as a threshold for distinguishing derived from non-derived pairs. It is important to mention here that resemblance values on different n-gram levels are not mutually comparable, and cannot be directly converted.

Results
General results
The results are presented in the Figure 2. As we see, F1-measure reaches its maximum on the level of 6-grams with the value of around 0.82. Results in more detail are given in Table 1.
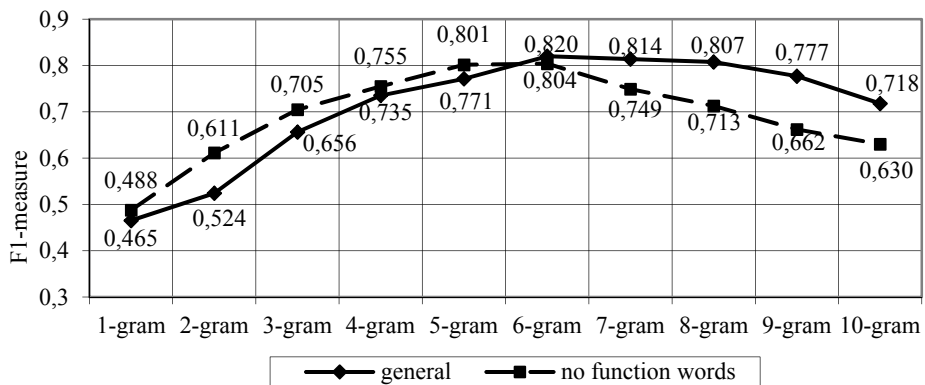
Figure 2: F1-measure maxima on the level of each n-gram respectively

Table 1: Detailed results of the analysis of measurements

	3-gram	4-gram	5-gram	6-gram	7-gram	8-gram
resemblance [%]	0.45	0.15	0.06	0.025	0.015	0.01
precision	0.603	0.717	0.755	0.833	0.868	0.894
recall	0.721	0.755	0.789	0.808	0.766	0.736
F1-measure	0.656	0.735	0.771	0.820	0.814	0.807

Removing function words
Function words are words that have little lexical meaning, but instead serve to express grammatical relations. These are regularly high-frequency words and most of them are common to all texts written in the same language.[15] Our inten-

---

[15] Manning and Schütze (1999, p.20)

tion was to reduce the number of n-gram overlaps caused by these high-frequency words. We assumed that by removing function words from the text we shall increase the sensitivity of the method. Words that will be treated as function words and removed from the documents in our text collection are 50 most frequent words from the Croatian National Corpus[16]. The results are also shown in Figure 2.

From the results we can see that even though F1-measure increased on the level of shorter n-grams, for longer n-grams we have significantly worse results. Surprisingly, overall result remained practically the same; measurement over original documents gave just slightly better score.

Distinguishing functional styles

As we know, language does not always function in the same way, but in that many ways that the society needs.[17] Differences can be found in various language features, such as lexical choice, syntactic constructions, lexical and sentential semantics, etc. And functional style is a subsystem of a language that comprises all features characteristic for a certain function. These differences can manifest themselves in a way that certain features are avoided in some styles while favored in others, or in some styles occur more frequently than in others. By distinguishing functional styles in the analysis, we wanted to check whether the functional style of the documents affects the results in some way.

According to Silić and Pranjković (2005, p.375), in modern standard Croatian language, five functional styles can be distinguished, and these are: scientific (*znanstveni*), official (*administrativno-poslovni*), newspaper and publicistic (*novinsko-publicistički*), literary (*književnoumjetnički*) and colloquial (*razgovorni*) functional style. Most of the documents from our text collection could be associated with one of the two functional styles, namely, scientific and publicistic and newspaper style. We have focused on these two styles for several reasons. Documents written in these styles were easily available and because of the ethical implications mentioned earlier, the question of originality is especially emphasized when speaking of texts written in these styles. Derivation pairs consisted of documents written in the same functional style were filtered and analyzed respectively. The results are shown in Figure 3.

The analysis showed that although considerable differences can be seen in the trend of F1-measure between the scientific and publicistic and newspaper style, overall results showed no significant change. It is interesting to notice, though, how for pairs written in the scientific style F1-measure remains practically con-

---

[16] These words are: *i, je, u, se, na, da, za, su, od, s, o, a, će, koji, ne, iz, što, bi, to, nije, ili, te, kako, kao, do, koje, biti, koja, godine, ali, samo, sve, jer, još, sam, više, po, sa, može, prema, već, nakon, dana, bio, bilo, zbog, li, smo, pa* and *ni*.

[17] Silić and Pranjković (2005, p.375)

stant from the 4-gram level to the level of 9-grams (difference is barely 0.02). On the other hand, pairs written in publicistic and newspaper style show the strong tendency towards 6-grams as indicators of derivation.
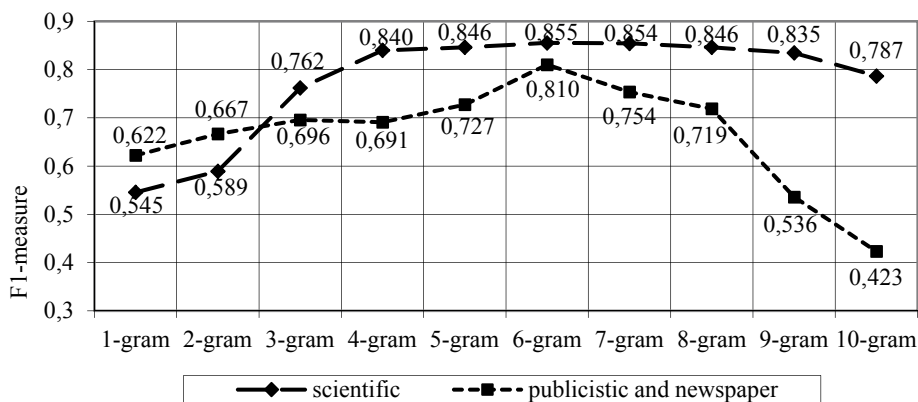

Figure 3: F1-measure maxima on the level of each n-gram respectively – differentiating functional styles

Conclusion

In this research, we have performed several measurements over the text collection to determine the most representative n-gram length for detecting reuse of text in Croatian language. We have taken 3 possible approaches to this problem: 1) measurements were performed over the original texts; 2) measurements were performed after the function words were removed from texts; and 3) derivation pairs consisted of documents written in the same functional style were analyzed separately. In every approach, we have proven that 6-gram is the optimal n-gram for detecting text reuse. Measurements by other authors on English texts have shown 7-gram to be the optimal one.[18]

Further on, we have concluded that removing function words as a text preprocessing method does not increase the sensitivity of the method, on the contrary, the quality of overall results decreases as n-grams are longer.

The first step in further research would be to enlarge the text collection and refine its system of text classification. The second would be to experiment with different kinds of text editing; e.g. POS tagging and extracting *hapax legomena*, stop words, labels or direct quotes. The third option would be to focus on a level of text other than words and observe it as a string of characters or sentences.[19]

Detection of derivation by direct translation from other languages is a challenge that goes beyond the scope of this paper, but should be very seriously consid-

---

[18] Clough (2003, p.185)

[19] As done in measurments on the METER corpus, Clough(2003, p.185)

ered when developing practical tools for use on documents in non-global languages.

We consider the results obtained by this research merely as a reference because the parameters which are to be used in a practical implementation highly depend on the requirements laid before it. Some implementations, for example in plagiarism detection systems, will probably favor recall over precision and thus choose shorter n-grams with lower resemblance threshold, while in others, such as identifying verbatim copies of news agencies articles, will most likely favor precision over recall and use longer n-grams with high resemblance threshold.

Nonetheless, we are confident that our findings will serve as a valuable contribution in developing methods and tools for automatic detection of derivation, which will find their application both in supporting academic integrity, and in aiding further research in the field of information sciences. The practical application of automatic methods for plagiarism detection should always be widely publicized so that it can help fight plagiarism by serving as a deterrent rather than a tool for enforcing justice and seeking punishment after the offense is already committed.

References

Bosanac, Siniša; Mandić, Bojana; Sprčić, Andrija. Objective Journalism or Copy-Pasted Press Releases: A Preliminary Media Content Analysis. // *INFuture2009: Digital Resources and Knowledge Sharing* / Stančić, H. *et al.* (ed.). Zagreb: Department of Information Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, 2009, pp. 417-425.

Bosanac, Siniša; Štefanec, Vanja. Preklapanje n-grama pri automatskoj detekciji deriviranosti dokumenata. // *StuLiKon – Studentska lingvistička konferencija*. Beograd: Univerzitet u Beogradu, 2011 (in print)

Clough, Paul D. Measuring text reuse. PhD thesis. Sheffield: University of Sheffield, 2003. Available at: ir.shef.ac.uk/cloughie/papers/thesis.pdf.

Clough, Paul D. Measuring text reuse and document derivation. PhD transfer report. Sheffield: University of Sheffield, 2001. Available at: ir.shef.ac.uk/cloughie/papers/transfer.pdf.

Manning, Christopher D.; Schütze, Hinrich. Foundations of statistical natural language processing. Cambridge; London: The MIT Press, 1999

Shivakumar, Narayanan; Garcia-Molina, Hector. SCAM: A Copy Detection Mechanism for Digital Documents. // *Proceedings of the Second Annual Conference on the Theory and Practice of Digital Libraries*. Austin, Texas, 1995. Available at: http://ilpubs.stanford.edu:8090/95/1/1995-28.pdf.

Silić, Josip; Pranjković, Ivo. Gramatika hrvatskoga jezika: za gimnazije i visoka učilišta. Zagreb: Školska knjiga, 2005

Wikipedia contributors. Jaccard index. *Wikipedia, the free encyclopedia*. May 23, 2011. http://en.wikipedia.org/w/index.php?title=Jaccard_index&oldid=430488441 (Accessed July 6, 2011)

Wise, Michael J. String Similarity via Greedy String Tiling and Running Karp−Rabin Matching. Sydney: Department of Computer Science, University of Sydney, 1993. Available at: www.pam1.bcs.uwa.edu.au/~michaelw/ftp/doc/RKR_GST.ps.

Štefanec, Vanja. Automatska detekcija plagijata. Unpublished paper. Zagreb: Faculty of Humanities and Social Sciences, University of Zagreb, 2010

Automatic Intonation Event Detection Using Tilt Model for Croatian Speech Synthesis

Lucia Načinović
Department of Informatics, University of Rijeka
Omladinska 14, Rijeka, Croatia
lnacinovic@inf.uniri.hr

Miran Pobar
Department of Informatics, University of Rijeka
Omladinska 14, Rijeka, Croatia
mpobar@inf.uniri.hr

Sanda Martinčić-Ipšić
Department of Informatics, University of Rijeka
Omladinska 14, Rijeka, Croatia
smarti@inf.uniri.hr

Ivo Ipšić
Department of Informatics, University of Rijeka
Omladinska 14, Rijeka, Croatia
ivoi@inf.uniri.hr

Summary

Text-to-speech systems convert text into speech. Synthesized speech without prosody sounds unnatural and monotonous. In order to sound natural, prosodic elements have to be implemented. The generation of prosodic elements directly from text is a rather demanding task. Our final goals are building a complete prosodic model for Croatian and implementing it into our TTS system. In this work, we present one of the steps in implementation of prosody into TTSs – detection of intonation events using Tilt intonation model. We propose a training procedure which is composed of several subtasks. First, we hand-labelled a set of utterances and within each of them, marked four types of prosodic events. Then we trained HMMs and used them to mark prosodic events on a larger set of utterances. We estimate parameters for each of the intonation event and generated f0 contours from the parameters. Finally, we evaluated the obtained f0 contours.

Key words: prosody in TTS, intonation model, Tilt

Introduction

Intonation modelling plays a great role in TTS systems. Synthesized speech without intonation component sounds unnatural and monotonous. Prediction of intonation patterns from text has been a difficult task due to their complex nature. There are, however, various prosodic models that predict prosodic elements from a text. They vary from rule based prescriptive models to data driven models such as CART decision trees (Dusterhoff et al., 1999), lazy learning approaches (Blin & Miclet, 2000) and unit selection based models (Meron, 2001). Phonological approaches to prosodic analysis of speech use a set of abstract phonological categories (tone, breaks etc.) and each category has its own linguistic function. An example of this approach is ToBI intonation model (Silverman et al., 1992). Parameter based approaches attempt to describe f0 contour using a set of continuous parameters. Such approaches are, for example, Tilt intonation model (Taylor, 2000) and Fujisaki model (Fujisaki & Ohno, 2005). Besides the mentioned models that tend to fall into one of the basic categories, there are models that use additional methodology (JEMA) (Rojc et al., 2005) or combine rule-based approach with data driven approach (Aylett et al., 2003). Regarding Croatian, there is a list of rules about how accents on words in a sentence are combined (Mikelić Preradović, 2008). Using method "analysis by synthesis", basic intonation categories: "rise", "fall" and "flat" have been determined (Bakran et al., 2001). Our goal is to build a complete prosodic model for Croatian and implement it into our TTS system. In this paper we will present the way we automatically detected intonation events for Croatian using Tilt intonation model and statistical models – hidden Markov models (HMM). In accordance with the results of the research on the basic intonation categories for Croatian, we have chosen Tilt intonation model which also differentiates three main prosodic events – rise, fall and connection.

The paper is organized as follows: in the next chapter we give an overview of the Tilt intonation model. In the third chapter we explain the procedure of the automatic detection of prosodic events. We describe the speech database we used, the process of hand-labelling, f0 feature sets extraction, the procedure of HMMs training and the process of f0 generation. We conclude the paper with the main results we obtained.

Tilt model overview

Tilt (Taylor, 2000) is a phonetic model of intonation that represents intonation as a sequence of continuously parameterised events (pitch accents or boundary tones). These parameters are called tilt parameters, determined directly from the f0 contour.

Basic units of a Tilt model are intonation events – the linguistically relevant parts of the f0 contour (circled parts in picture 1). From such a representation, it is possible to encode the linguistically relevant information in an f0 contour, and then recreate the original f0 from this coding.
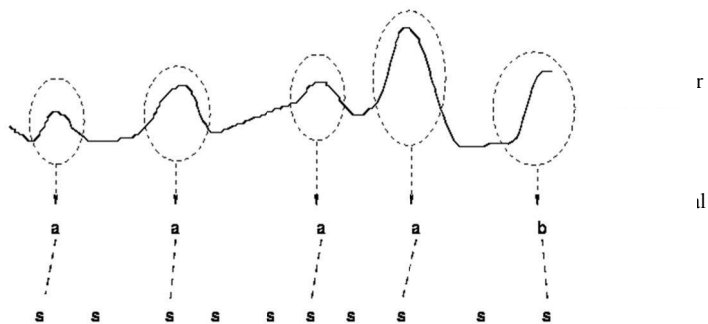
384

Figure 1: Intonation events in the Tilt model

Tilt model can be described with a simpler model – RFC model (R-rise, F-fall, C-connection). In the RFC model, each event is modelled by a rise part followed by a fall part. Each part has an amplitude and duration, and two parameters are used to give the time position of the event in the utterance and the f0 height of the event. (Taylor, 1995).
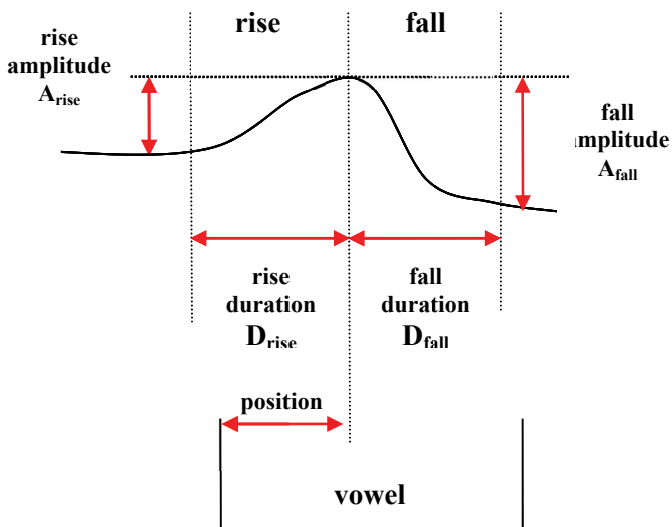

Figure 2: RFC parameters in the Tilt model

The RFC parameters for an utterance are:
- rise amplitude (Hz),
- rise duration (seconds),
- fall amplitude (Hz),
- fall duration (seconds),
- position (seconds),
- f0 height (Hz).

Those parameters can be transformed into Tilt parameters:
- Tilt-amplitude (Hz): the sum of the magnitudes of the rise and fall amplitudes:

$$tilt_{amp} = \frac{|A_{rise}| - |A_{fall}|}{|A_{rise}| + |A_{fall}|}$$

- Tilt-duration (seconds): the sum of the rise and fall durations:

$$tilt_{dur} = \frac{|D_{rise}| - |D_{fall}|}{|D_{rise}| + |D_{fall}|}$$

- Tilt: a dimensionless number which expresses the overall shape of the event, independent of its amplitude or duration:

$$tilt = \frac{|A_{rise}| - |A_{fall}|}{2|A_{rise}| + |A_{fall}|} + \frac{|D_{rise}| - |D_{fall}|}{2|D_{rise}| + |D_{fall}|}$$

Tilt is calculated from the relative sizes of the rise and fall components in the event. A value of +1 indicates the event is purely a rise, -1 indicates it is purely a fall. Any value between says that the event has both a rise and fall component, with a value of 0 indicating they are the same size.

Intonation event detection
In order to detect intonation events and label the whole database, an automatic HMM based procedure which we described in this chapter was used. The procedure uses four HMMs to predict the four intonation events from the f0 features. To train the parameters of the HMMs, a set of hand labelled utterances was used.

Speech database
Speech database that we used in our research consists of 1 hour and 54 minutes of speech from a collection of fairy tales spoken by one speaker. We hand-la-

belled hundred utterances out of which we selected a subset of 25 utterances for testing and 75 for training the model.

Hand labelling

From the speech database, we chose a hundred of utterances which were labelled by hand to produce intonation transcriptions. We located pitch accents, boundaries, connections and silences within each utterance, in accordance with the intonation event model (Figure 3.)
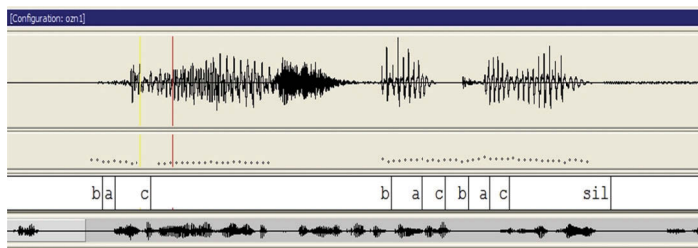

Figure 3: Intonation event transcription of an utterance

Labels that we used for labelling events are: *sil* for unvoiced parts, *a* for pitch accents, *b* for all rising boundaries and *c* for all falling boundaries.

The process of labelling was performed using the WaveSurfer tool (Sjölander & Beskow, 2000) . Labelled files were exported and saved as xlabel files and were then used in the automatic detection of intonation events.

F0 feature sets extraction

To extract f0 features from the training set of utterances we used RAPT algorithm (Talkin, 1995) as implemented in Voicebox Matlab toolbox. The f0 was sampled at 10 ms. The obtained f0 contours contained some noise which we smoothed with a three point median filter. We set the f0 value to 0 Hz to represent the unvoiced segments where f0 cannot be determined and in another attempt we used linear interpolation to determine the missing values. We got three different f0 feature sets: raw output from the RAPT algorithm, smoothed and interpolated. In addition to the f0 contour, we used dynamic features (delta f0 and delta-delta f0) for training HMMs of intonation events.

Automatic event recognition

We trained HMMs for detection of accents, boundaries, connections and silences. A five-state HMM was used for each event type. Models were trained with Baum-Welch algorithm on f0 features and hand-labelled event positions.

For all three f0 feature variants mentioned in the chapter above, a different HMM set was trained.

387

To limit valid event sequences, a grammar with permitted combination of events was defined. Viterbi algorithm was used to detect the events.

For testing the automatic event detection, the utterances are divided into two sets which were identical to the sets that we used in the process of training. We applied models trained on different f0 features from the training set to the set of utterances with different f0 features from the test set. The performances for all type of events and for each event separately are shown in Table 1.

Table 1: Performance for different feature sets

Feature set	Correctness
f0 raw	45.62
f0 smoothed	53.75
f0 interpolated	45.77

The correctness was computed using the Levenshtein distance between the automatically generated and hand-labelled event labels. We got the best results with models trained on median filter smoothed f0 feature set and applied to feature set obtained in the same way.

The interpolation of missing f0 values did not improve the event detection, as distinguishing between voiced and unvoiced speech may give important clues for event locations, and by interpolation this information was lost.

Tilt analysis

When the events are detected, the location of the start, peak and end position of each event has to be determined. The analysis performs only on those parts of f0 contours which were detected as the events. Each of those parts is smoothed by median smoothing algorithm and unvoiced regions are interpolated. Each event has to be described as a rise or fall shape within the f0 contour so tilt parameters have to be assigned to each of them. Algorithm used in tilt analysis minimizes the difference between the original contour and the fitted shape. We get a model represented by tilt parameters which were explained in chapter 2.

Tilt synthesis

From tilt/RFC parameters, we can generate f0 contours using tilt synthesis and given equations:

$$f0(t) = A_{abs} + A - 2.A.(t/D)^2 \qquad 0 < t < D/2$$
$$f0(t) = A_{abs} + 2.A.(1 - t/D)^2 \qquad D/2 < t < D$$

where A is rise or fall amplitude, D is rise or fall duration and A_{abs} is the absolute f0 value at the start of the rise or fall, which is given by the end value of the previous event of connection.

Places on the f0 contour between the events are filled using the method of interpolation.

Results

The usual measure for evaluating generated f0 contour is the root mean square error (RMSE) between the original contour and the obtained generated f0 contour. We compared the performance of three models for automatic event detection, trained on raw, smoothed and interpolated f0 features. The models produced event labels for the test data set (f0 features extracted from 25 utterances, with interpolation for unvoiced segments). Tilt analysis was performed using these labels and f0 features, yielding tilt parameters from which the f0 contours were synthesized. The resulting f0 contours were compared with the original (interpolated) f0 contour. In the same way the f0 contour was synthesized using hand-labelled events and compared with the original f0. The results of comparison are shown in Table 2.

Table 2. Mean RMSE values for generated f0 contours.

Event label model	RMSE (Hz)
raw	**25.16**
smoothed	26.69
interpolated	25.57
hand-labelled	**23.11**

We got the best results using the model trained on raw (unprocessed) f0 features. The obtained results are satisfactory but further improvements might be achieved.

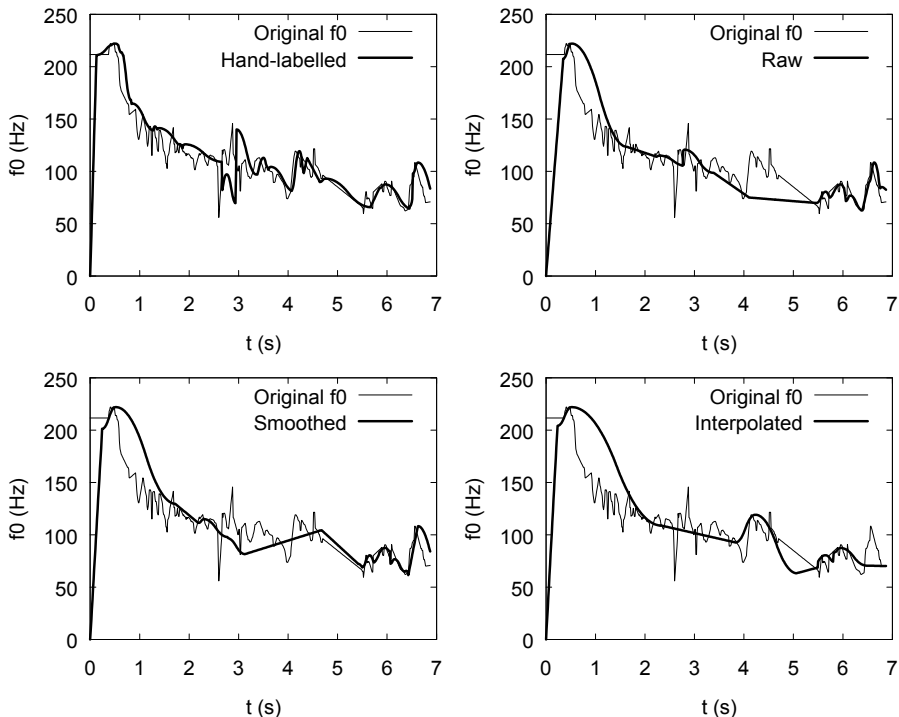Figure 4 shows an example of the generated contours, each compared with the original f0.

Figure 4: Comparison of the generated f0 contours with the original f0

Discussion

All f0 contours obtained from automatically detected events have similar RMSE values, and perform comparably to the hand-labelled case. More hand labelled training data may not be sufficient to improve the RMSE, but also improving the hand-labelling procedure could contribute to better RMSE.

We plan to improve the quality of hand-labelled event boundaries using an automated procedure. A search for the optimal position of the boundary could be done by trying several positions in the vicinity of labelled boundary and noting the change in observed RMSE. The boundary is fixed after a predefined number of iterations.

Further step towards automatic f0 generation from text will be CART (classification and regression trees) building. Based on the questions in tree nodes regarding chosen linguistic features extracted from text, trees will predict tilt parameters.

Conclusion

Implementation of prosodic elements into text-to-speech systems represents a demanding task. As a part of our final goal to implement prosody into our TTS, in this paper we proposed a procedure for automatic event detection. We chose a representative set of utterances and marked four main prosodic events within each utterance. We trained HMMs to mark events automatically on a larger set of utterances. We parameterized the detected events with tilt parameters and generated f0 contours out of those parameters. We evaluated the obtained f0 contours. Future work will include building a model for prediction of tilt parameters from text.

References

Aylett, M.P.; Fackrell, J.; Rutten, P. My voice your prosody: Sharing a speaker specific prosody model across speakers in unitselection tts. // *Eurospeech*. 2003; 321-324.

Bakran, J.; Erdeljac, V.; Lazić, N. Modeliranje temeljnih intonacijskih oblika. // *Govor*. vol. 18 (2001); 105-111.

Blin, L.; Miclet, L. Generating synthetic speech prosody with lazy learning in tree structures. // *CoNLL-2000 and LLL-2000*. 2000; 87–90.

Dusterhoff, K.E.; Black, A.W.; Taylor P. Using decision trees within the tilt intonation model to predict f0 contours. // *Eurospeech*. 1999; 1627–1630.

Fujisaki, H.; Ohno. Analysis and modeling of fundamental frequency contours of English utterances. //*Speech Communication*. 2005; 47:59-70.

Meron, J. Prosodic unit selection using an imitation speech database. // *4th ISCA Workshop on Speech Synthesis*. 2001; 53-57.

Mikelić Preradović N. Pristupi izradi strojnog tezaurusa za hrvatski jezik. Ph.D. thesis, University of Zagreb. 2008.

Rojc, M.; Agüero, P.D.; Bonafonte, A.; Kacic, Z. Training the Tilt intonation model using the JEMA methodology. // *Eurospeech*. 2005; 3273-3276.

Silverman et al. ToBI: A standard for labeling English prosody. // *ICSLP92*. 1995; 2:867-870.

Sjölander, K.; Beskow, J. Wavesurfer – An open source speech tool. // *Interspeech.* 2000; 464-467.

Talkin, D. A robust algorithm for pitch tracking (RAPT). // *Spech coding and synthesis*. 1995; 495-518.

Taylor, P. Analysis and synthesis of intonation using the tilt model. // *The Journal of the Acoustical Society of America*.2000; 1697-1714.

Taylor, P. 1995. The rise/fall/connection model of intonation. // *Speech Communication*. vol. 15 (1995); 169–186.

DIGITAL PRESERVATION

Croatian Digital Library Initiatives

Dunja Seiter Šverko
National and University Library
Ul. Hrvatske bratske zajednice 4, Zagreb, Croatia
dseiter-sverko@nsk.hr

In response to the needs of the users in digital age the National and University Library in Zagreb has been developing projects and services for the foundation of the Croatian digital library. The basic Library's roles to identify, acquire, collect, store, organise and provide long-term access to the national heritage are adapted to suit the rapidly developing networked environment. The Library's digital collections include born digital resources - web publications and digital masters, as well digital reproductions of different library materials. The presentation provides an overview of some current projects and activities: Digitised Heritage – a web site presenting a selection of digitised rare and unique works from Library's special collections. The most valuable works of the Croatian written heritage have been digitized, such as the first Croatian printed book from 1483, the works of old masters and contemporary artists, along with maps, music prints, posters and photographs. DAR – Digital academic repository is a portal for search and retrieval of electronic theses and dissertations of the University of Zagreb. The system serves as an OAI-PMH compatible repository for digitized research theses (1880-1952) and digital born theses (from 1986 on). The portal provides a bio-bibliographic reference data base on more than 800 authors created in cooperation with the Miroslav Krleža Institute of Lexicography. DNC – Digitized newspapers and journals is a cooperative portal of the National and University Library in Zagreb and other heritage institutions that provides access to digitized Croatian historical newspapers and journals. DNC is the central point for search and access to digitized serials that aims at improving preservation of unique and rare originals, enhancing development of standardised procedures for digitization, and enabling coordination and rationalisation of digitization efforts. HAW – Croatian Web Archive is a system for gathering and storage of the legal deposit copies of Croatian web resources with scientifically or culturally relevant content. The archived content represents part of the Croatian national heritage, and supplements the national collection. The focus is on the born digital content existing on the web only, which documents the stuff of everyday life, social trends, popular events, important sport, political, cultural, or other types of events. The implementation of Aleph Integrated Library System in the National and University Library in Zagreb, and the libraries of the University of Zagreb and public scientific institutes, outlines the foundation of Croatian integrated library system (ILS). ILS comprises the holdings of forty four libraries, with a total of 3,000,000 volumes, and 140,000

library users. The objectives of ILS implementation were to rationalize work processes, to raise the quality of service and to foster universal access while securing autonomy and ownership of individual library holdings. In addition to numerous advantages of digital library services including easier and faster access to books, periodicals, maps and other library materials of interest to scientists, professionals, students and general public, the realization of these projects provides the basis for the implementation of best practices in digitization, application of international standards and creation of national policy for the collection, storage and access to digital materials. The ILS provides the general framework for cataloguing and metadata creation and serves as a uniform interface to different digital collections of the National and University Library in Zagreb.

Usage of Virtualization Technologies in Long-Term Preservation of Integrity and Accessibility of Digital Data

Ivor Milošević
University Computing Center
Marohnićeva 5, Zagreb, Croatia
ivor@srce.hr

Hrvoje Stančić
Department of Information Sciences,
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
hrvoje.stancic@zg.t-com.hr

Summary

The concept of virtualization making a transition from expensive hardware systems over to the affordable software field, capabilities that modern virtualization environments allow are becoming multidimensional and not many software designers can allow their applications not to include some type of virtualization system compatibility. Many applications allow virtualization systems to utilize them in such a manner that the performance is increased rather than hindered, and the level of collaboration among the users to be increased. Distribution of information, resources as well as entire operating environments can be very successfully and appropriately applied and used in systems for maintaining authenticity and preserving long term usability of data. The authors analyse requirements and suggest some initial high-level architectural solutions for the users of virtualization in the long term preservation process.

Key words: data preservation, virtualization

Introduction

Archival institutions, as well as all institutions requiring long-term preservation of digital data, whether in the form of documents, images, sound, moving images or 3D objects, are faced with the problem of hardware and software obsolescence. Those institutions, more often than before, have large scale datacenters implemented as the supporting facilities for their mandate operations. Accessibility and integrity of the data they preserve are obligatory requirements, sometimes followed by the requirements for authenticity, reliability and usabil-

ity.[1] Although those requirements may seem easy to fulfill, it is not the case if long-term preservation perspective is taken into account. During the next 10, 30, 50 or more years of preservation of digital records, the underlying technology, upon which relies the possibility of fulfilling the mentioned set of requirements, will have to be changed several times. The question is how to achieve that as seamlessly as possible? The best solution would be that one can change the underlying technology, i.e. the lower layer, without affecting the data on the top layer. The technology which is far from a perfect preservation technology, yet offering good results, is the virtualization technology. The idea is to use virtual machines, i.e. software solutions for creation of virtual (software-based) computer(s) on one physical, modern computer. This would allow data-preserving institutions to change the actual hardware as needed to keep up with recent developments still having the obsolete systems installed and working on a virtual machine.

Emulation, to which virtualization is in some way similar, is an approach which "does not focus on the digital object, but on the hardware and software environment in which the object is rendered. It aims at (re)creating an environment in which the digital object can be rendered in its authentic form"[2]. In that sense emulators are mainly software solutions that can be installed on the current operating systems and emulate older hardware and/or software. On the positive side it is possible to run obsolete programs or read data in obsolete formats on the current hardware and software, but on the negative side emulators are written for specific hardware and software systems and can become obsolete too. Contrary to this, virtualization on the positive side:

- eliminates the influence of system upgrades on the preserved data,
- eliminates the compatibility problems of running old software on the new hardware,
- makes possible to install more than one virtual machine on a single physical machine thus enabling the possibility of having a parallel networked virtual computers environment,
- makes possible to export data from the preserved data format to other, possibly more stable, format or format which was supported by the original application and which has proved, during a period of preservation time, easier to migrate to a newer format. Such capabilities enhance the use of virtualized solutions in the environments where hybrid preservation approaches are applied.

---

[1] As in: *ISO 15489-1 – Information and documentation – Records management*, 2001., 7.2.2., http://www.iso.org.

[2] Hoeven, Jeffrey van der, Wijngaarden, Hilde van, *Modular emulation as a long-term preservation strategy for digital objects*, 5th International Web Archiving Workshop (IWAW05), 23-23 September 2005, Vieanna, http://iwaw.europarchive.org/05/papers/iwaw05-hoeven.pdf (01. 09. 2011)

On the negative side this approach has similar drawbacks as the approach of maintaining original technology:

- requires trained personnel for operating obsolete systems and helping users
- as Thibodeau (2002) says, having in mind a wider problem of digital preservation, "it would cut users off from the possibility of using more advanced technologies for discovery, delivery, and analysis".

Why virtualization?

Modern datacenters and information technology systems are becoming increasingly reliant on techniques of masking physical characteristics of computing resources from the way in which other systems, applications or users access and utilize those resources. This includes a form of morphing of one physical resource (server, operating system, application or storage systems) into apparently functioning as multiple logical resources or vice versa – making physical resources appear and function as a single logical one.

Virtualization can therefore be defined as an abstraction of computing resources with the goal of simpler, more efficient and more thorough usage of information systems hardware as well as providing advanced functionalities such as high availability, automated resource distribution and organization etc.

Such virtualization systems, besides being used for server consolidation, simpler datacenter management etc., present a solid foundation for long-term data preservation systems although this is not their intended or primary purpose.

With adequate planning and appropriately defined information object for long term preservation, any of the currently available virtualization systems and infrastructures with adequate hardware background can be successfully adapted for this purpose.

Virtualization systems in this model are only indirectly capable of preserving individual files and formats (e.g. pdf, doc, xls, zip, mp3) via "information blocks" that can include complete operating environments, virtual disks, data storage systems and similar, which are then directly accessed by the users and which offer the ability to not only access and view the data but also provide the tools necessary for their usage and manipulation.

Hardware requirements

Initially it is necessary to define the hardware requirements in which such a model can operate. Let's consider a virtual machine on a standalone virtual host. In this case, the virtual machine with all of its accompanying virtual disks and data is the information block which is intended for preservation.

Virtualization host should be equipped with redundant hard drives in order to ensure security of data in a hardware failure scenario as well as multiple failo-

ver paths to the network interfaces and the storage system. All the connections should be interlaced as shown in Picture 1.

Such an approach allows both high availability schemes to be implemented, as well as safety mechanisms from data corruption due to hardware malfunction.
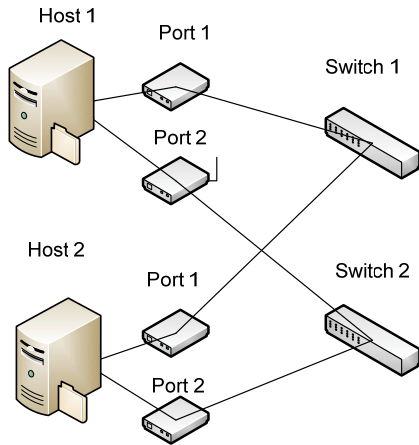

A standalone virtualization host implemented in this manner allows for the identically configured host to be added into the system, therefore allowing instantaneous transfer of all operations, managed by the virtualization software, in case of a catastrophic hardware malfunction. In order to expand the capabilities further, both hosts must have access to a central storage repository which in itself provides a hardware level redundancy on the virtual disk layer.

Picture 1: Redundancy and failover paths

All the data of the aforementioned information block, virtual machine or disk, with all the accompanying data should, in this case, be stored on the shared storage system while the virtual hosts' internal disks should be used only for virtualization software requirements. Ability of the system to transfer the entire running operating environment from one physical hardware to another in real-time can be further expanded using the true high availability systems. They allow uninterrupted sharing of active data, thus creating a clone of the virtual environment capable of uninterrupted "takeover" of shared data access and processing.

Since recommendations regarding virtualization systems are stating that the load of each individual host within the virtualization system and all of its components should optimally be under 50% (due to the failover requirements), now that a clone of the original system has been created, a fully clustered system needs to be established. This brings the minimum required number of virtualization hosts per cluster to 4, including a shared storage system.

Implementation of any of the virtualization tools on the specified hardware and addition of a remote safety copy of the virtualization component as well as replication of the storage component is shown as a system diagram in Picture 2.
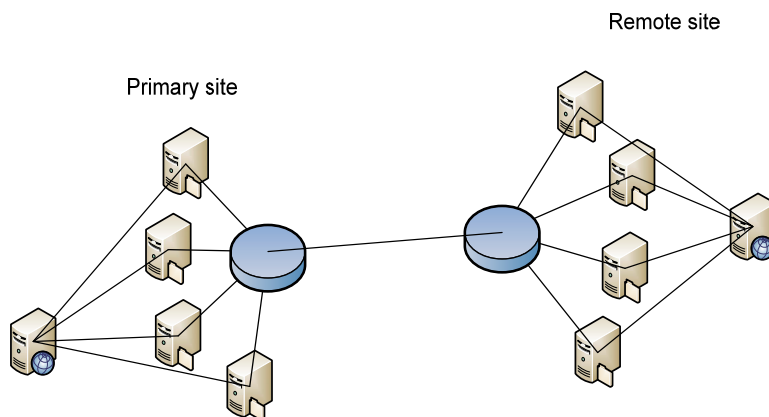
Picture 2: Fully redundant virtualization cluster with replicated storage

The key here is replication of the storage system so that, in an emergency situation, the remote location can independently take over all the operating components of the primary.

Components of virtualization systems

Virtualization systems are a collection of tools used in wide range virtualization of the information technology infrastructure which allow for a variety of advanced features in order to ensure high availability, automated resource management, monitoring, distribution and control over data usage within the system as well as many other functions.

Several virtualization tools (hypervisors) available today offer the required functionality:

1. VMWare – vSphere (ESX, ESXi), vmotion abilities in the commercial variant
2. Citrix – Xenserver 5.5+ freeware version
3. Microsoft – Hyper-V in MS Windows domain environment

The data object can now be defined as any format which can be interpreted, read, modified or acted upon in any manner allowed by the virtualization system functionalities. This can include:

- an entire virtual machine including operating system and/or desktop working environment,
- virtual disk in a format supported by the relevant hypervisor,
- data on the virtual disk via direct access to the storage system.

The entire virtual disk can be placed in an inactive state on the specified hardware, where, due to the multiple failover redundancy of each component, a practical indestructibility of stored data is ensured. Template or clone can con-

401

tain and store such a "block", or it can be copied onto a storage system which prevents unwanted or unauthorized copying. Such a "safety" copy allows both verification of its integrity by comparing it to the identical data stored on the backup system and incorporation into a versioning system schema, where all the potential modifications are visible and directly tied to a specific user.

The most important aspect of such a model is the creation of a "infoterminal" type system where the virtual desktop environment, as well as all the data which is intended for long-term preservation, is stored with the tools and instructions necessary for viewing and operating such data. All of those elements are then grouped together into a single un-modifiable environment which utilizes all of the previously listed advantages of the virtualization cluster. In this way, the user is presented with a complete virtual interface, with all the necessary tools and physically (hardware-level) enforced integrity of the data within the central repository. The central repository allows the virtual environment to be copied but it subsequently manipulates and process the data on another location without compromising the original.

Due to the fact that the entire system is operating in a virtual environment, the number of users is theoretically indefinitely scalable. In case the number of users of a certain virtual "infosystem" is increased and the resource requirements exceed the resources allocated to a service in question, virtual environment allows for simple and even automatic resource redistribution without adverse effects on the guest virtual machine. Also, if there is a need to upgrade or format the existing system, virtualization makes it very simple to allow a parallel cohabitation of two systems. Therefore, usage of the old one can remain uninterrupted during migration to the new one. It is also possible to retain all the old formats and tools required to access legacy data in case the format migration process entails a loss of certain characteristics (e.g. forward migration from doc to docx).

Such a system, based on any of the currently available hypervisors, safeguards the data from the need of version changes or the format of the data within the virtual PC itself, since such a machine is independent from both hypervisor or hardware upgrades or any other external factors. So what is obtained is simultaneous preservation of data as well as tools required for manipulation of that data.

The user is now able to access the desired virtual environment from any location, where network access is granted to the virtualization management tools, and use personal access data through the necessary client software available for various operating environments. The background of the virtualization cloud (Picture 3) and its internal modifications, upgrades etc. do not affect the end user in any way.
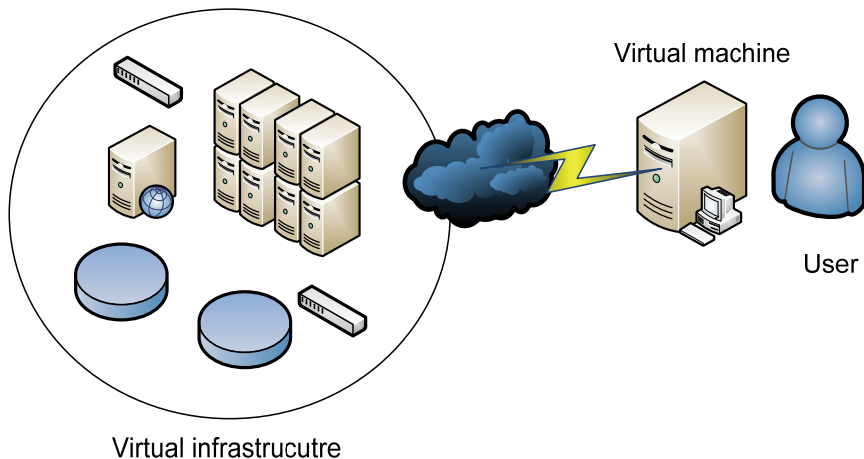
Picture 3: End-user view of the virtual infrastructure

Therefore we suggest usage of virtualized systems because of at least eight situations concerning the long-term preservation of integrity and accessibility of the data:

1. Hypervisor upgrades do not affect virtualized data or machines in any way, which ensures longevity of the data in case when virtualized systems need upgrades.
2. Modern storage systems, with implied regular maintenance, practically guarantee a 100% safety from data loss or corruption in long term-data preservation due to the fact that the data is not affected in the case of hardware component malfunction. Additionally, hardware can be upgraded and replaced without either disrupting normal operation of the entire system or the integrity of the data it preserves.
3. If upgrades of the storage system are required, hypervisors offer the ability for data migration from one storage unit to another while maintaining data integrity and real-time functionality of the entire infrastructure.
4. Incorporation of the solid state drives (SSD) in the system, which possess no movable components, will further increase safety of the stored data.
5. Virtualization methods of creating backup copies on the virtual disk layer, or having the entire dataset stored in the virtual machine offer a redundant mechanism for data loss prevention, thus fortifying the preservation process.
6. Hypervisors offer the ability of storing as well as using old formats and systems without the need of software upgrades and without adverse effects of hardware upgrades, while maintaining a low cost of upkeep for old or obsolete technologies.

403

7. Backwards compatibility is ensured because new versions of the hypervisor implicitly support all of the previously supported variants of operating systems.

8. The ability to create templates and store the entire operating environments as well as to ensure their integrity through authorization and authentication mechanisms of virtualization systems adds to the stability of the long-term preservation process.

It can appear that the described systems are expensive or inaccessible to smaller or financially weaker institutions. However, many modern datacenters are already, in one way or the other, moving towards being virtualization compatible, for the reasons of maintenance cost reduction as well as better utilization of existing resources and infrastructure, many of which can support the desired virtualization tools in freeware versions, therefore incurring no additional cost.

Furthermore, due to the extensive flexibility of such systems, with an adequate network infrastructure, it is possible to share resources between institutions or to host external services, thus making it unnecessary for all institutions to own all the components of the discussed model. Instead, the resources can be shared for financial compensation, decrease of the maintenance costs, or by some other arrangement. This is in accordance with the OAIS RM suggestion for technical level of interaction between OAIS archives, i.e. for creation of archives with shared functional areas. This means that "each archive can serve totally independent communities (...). However, for the common storage element to succeed, standards are needed at the internal Ingest-storage and Access-storage interfaces."[3]

Interchange of materials, as well as complete viewers of data between institutions is in this way reduced to the procedure of copying the virtual disk, or merely allowing access to the intended user.

Conclusion

Successful long-term preservation of digital data depends on the ability to safely store, maintain integrity over extended periods of time, and ultimately allow changes of underlying technology while retaining accessibility of the digital data which is being stored and preserved. With recent advances in virtualization technologies, tools have been developed that allow all of these goals to be achieved in relatively small or low-cost environments depending on the size and scope of the digital data in question.

Although primarily targeted at allowing the hardware consolidation, creation of affordable high-availability systems as well as easier-to-administer datacenter system, virtualization has brought with it the ability to logically detach com-

---

[3] *Rerefence Model for an Open Archival Information System (OAIS)*, Blue Book (CCSDS 650.0-B-1), Consultative Committee for Space Data Systems, NASA, Washington, DC, SAD, January 2002, pp. 6-7, http://public.ccsds.org/publications/archive/650x0b1.pdf (15. 03. 2011)

plete "self-sustained" software environments, which include operating systems, applications and the digital data of any format or type, from the hardware and/or software on which it depends.

Therefore we are suggesting the usage of the virtualization systems, and all their advanced functionalities, to collectively save the digital data intended for preservation, their operating environment, and all the tools needed for access, intended and allowed modification and usage of the data. The data will stay preserved regardless of the future changes of format, type, software capabilities and compatibilities as well as other problems present in the long-term preservation of digital data.

Furthermore, implicit high-availability and storage functionality of the virtualization systems both safeguards data integrity and ensures accessibility regardless of hardware failures and underlying technology modifications, upgrades and/or adjustments. Considering the fact that virtualization systems are increasingly becoming a common component of medium-scale or even small-scale datacenters, they are a logical and effective way of applying the existing technology mechanisms in achieving, often complex, long-term data preservation goals.

Further research

Taking into account long-term preservation of digital data, virtualization techniques prove to be good solution for the period of time longer than the expected life-span of the standard digital preservation media. Virtualization will prove efficient for the data frozen in time and for testing and making migration procedures easier and more transparent. Nevertheless, there are some data, e.g. results of certain experiments, measurements etc., that will need to be used for future analysis. In some cases emulators will be needed in order to achieve the original functionality since standards and protocols will eventually change. Further research is planned on the role and possibilities of using emulators along with virtualized systems for extension of the periods during the long-term preservation process in which both the preserving institutions and users could be sure that the integrity of the data is preserved and that the data are accessible, usable and reliable.

References

Hoeven, Jeffrey van der, Wijngaarden, Hilde van, *Modular emulation as a long-term preservation strategy for digital objects*, 5th International Web Archiving Workshop (IWAW05), 23-23 September 2005, Vieanna, http://iwaw.europarchive.org/05/papers/iwaw05-hoeven.pdf (01. 09. 2011)

*ISO 14721:2003, Open Archival Information System – Reference Mode*l, Space Data and Information Transfer Systems, http://www.iso.org

ISO 15489-1 – Information and documentation – Records management, 2001. http://www.iso.org

Kemp, Yves, *Use of Virtualization Techniques for Data Preservation and Long Term Analysis*, First Workshop on Data Preservation and Long Term Analysis in High Energy Physics, Janu-

ary 26-28, 2009, DESY, Hamburg, Germany,
http://www.dphep.org/sites/site_dphep/content/e20/e36625/e43750/dlpta.pdf (07. 07. 2011)
Rerefence Model for an Open Archival Information System (OAIS), Blue Book (CCSDS 650.0-B-1), Consultative Committee for Space Data Systems, NASA, Washington, DC, SAD, January 2002, http://public.ccsds.org/publications/archive/650x0b1.pdf (15. 03. 2011)
Smith, David, *The Bleeding Edge: Virtualization*. 18 February 2008,
http://www.gantthead.com/content/articles/240292.cfm (22. 01. 2011)
Thibodeau, Kenneth, Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years, in: *The State of Digital Preservation: An International Perspective*, Council on Library and Information Resources (CLIR), Washington, D.C., SAD, July 2002., pp. 4-31, http://www.clir.org/pubs/reports/pub107/pub107.pdf (07. 07. 2011)
VSphere Availability Guide. http://www.vmware.com/pdf/vsphere4/r40/vsp_40_availability.pdf (15. 04. 2011)

Supporting e-Science:
Scientific Research Data Curation

Radovan Vrana
Department of Information Sciences,
Faculty of Humanities and Social Sciences, University of Zagreb,
Ivana Lučića 3, Zagreb, Croatia
rvrana@ffzg.hr

Summary

One of the outcomes of the continuous development of science is creation of new methods of scientific research which, as a result, generate different types of research output including research data. While a significant attention is given to preservation of journal articles, books, and papers published in conference proceeding, less attention is given to the preservation of research data. To enable use of data accumulated in previous scientific research projects in a new scientific research, research data should be preserved. The activity of research data preservation is called data curation. Data curation has become necessary if science wants to avoid data loss. Unfortunately, science itself cannot take care of research data easily; it needs help from professionals like librarians and archivists to preserve research data in order to enable their re-use future scientific research. Although there are already some good solutions to this problem such as storing research data in digital repositories, no final decision has been made about who will take the responsibility for this kind of activity in the long run.

Keywords: scientific communication, e-science, data curation

Introduction

One of the outcomes of the continuous development of science is a creation of new methods of scientific research which, as a result, generate different types of research output including research data. While usual scientific output in form of journals articles, books, and conference papers is preserved and made available in form of local digital collections in libraries owned by academic institutions or in remote full text databases for access to which academic institutions are paying licenses, research data is rarely preserved. The research data have a special value, since they can be included into new scientific research projects and lead to new scientific discoveries. To enable such use of data accumulated in previous scientific research projects, research data should be preserved. As this

problem of preservation of research data attracted more attention over the years, scientific community in cooperation with libraries and archives started to consider development of a support in form of procedures, policies, guidelines and standards for long term preservation of research data in various formats. In addition to these written documents, a new activity of long term preservation of research data was initiated and it became known as data curation. The purpose of this activity is to facilitate re-use of research data created as output of a previously completed scientific research in a new scientific research. Inclusion of research data in the new research has become possible because of the changing role of data in the scientific world. Research datasets have ceased to be merely the output of the research endeavour, and they have become a new input to new hypotheses which enable new scientific insights and drive innovation (National Science Foundation, 2007). Consequently, it has become necessary to enable systematic capturing and preservation of the scholarly output with a special attention given to the scientific research datasets in order to minimise the risk of data loss in an ever-changing environment "(...) where data flows and technologies are changing constantly." (Angevaare, 2009, p. 4). The risk of data loss is becoming greater as our reliance on digital information resources grows as well as our negligence about the future of the research output in digital form: "By the time knowledge in digital form makes its way to a safe and sustainable repository, it may be unreadable, corrupted, erased, or otherwise impossible to recover and use." (Ogburn, 2010, p. 242). Ogburn (2010, p. 242) offered reasons why research data may be endangered: "(…) due to their sheer size, computational elements, reliance on and integration with software, associated visualizations, few or competing standards, distributed ownership, dispersed storage, inaccessibility, lack of documented provenance, complex and dynamic nature, and the concomitant need for a specialized knowledge base—and experience—to handle data." Data curation has also become necessary because the scientific endeavour has become too expensive to let research data be destroyed or forgotten. Some research data are unique and cannot be recreated which make them primary concern for their long term preservation. Science itself cannot take care of research data easily; it needs help from professionals like librarians and archivists to preserve research data in order to enable their re-use future scientific research. This paper will give an overview of the problem of data curation and emphasize its significance to the modern science.

E-science, research data and data curation

Nowadays, when we speak about the modern science, we usually refer to ICT supported networked science that is global, data intensive and more collaborative than ever. We call such science - e-science. According to Lord and MacDonald (2003, p. 5) "The term e-Science – or more inclusively e-Research - has been used recently to describe the research culture and opportunities enabled by

these developments, and the collaborations of people and of shared resources that are needed to resolve new research challenges, whether in the sciences, social sciences or humanities. E-science enables a new order of collaborative, more inter-disciplinary research, based on shared research expertise, instruments and computing resources, and crucially increasing access to collections of primary research data and information - the knowledge base of research.". E-science is characterized by a creation of great quantities of data "(…) generated from sensors, satellites, high-performance computer simulations, high-throughput devices, scientific images and so on (…)", and they will "(…) soon dwarf all of the scientific data collected in the whole history of scientific exploration." (Hey and Trefethen, 2003, p. 4). To facilitate global collaboration, e-science needs an infrastructure that will make possible "(…) sharing of computing resources, data resources and experimental facilities in a much more routine and secure fashion than is possible at present." (Hey and Trefethen, 2003, p. 1). Digital technology should provide such infrastructure and secure long term preservation of "(…) data generated today so it can survive the changes of technology and can be accessed in the future." (Hockx-Yu, 2006, p. 234). Taking care of the research data in the long term will be a very difficult task as it requires complete understanding of versatile and complex digital objects, procedures which create these data, procedures to decide what data to keep, hardware and software necessary to complete data care, trained staff that will operate digital repositories of research output etc.

Different scientific fields have different types of research output. When speaking about the long term preservation in science (in general), one usually refers to the preservation of printed scientific output such as journal articles, books, and conference papers published in proceeding but also to research data. The term data is here used "(…) to refer to any information that can be stored in digital form, including text, numbers, images, video or movies, audio, software, algorithms, equations, animations, models, simulations, etc. Such data may be generated by various means including observation, computation, or experiment." (National Science Foundation, 2005, p. 9). For Beagrie, Chruszcz, and Lavoie (2008, p. 18) "Research data is an essential input to scholarly endeavour, whether that endeavour is focused on extending the frontiers of knowledge, or understanding the discoveries of the past. (…)". To make use of such research data, e-science needs a well organized middleware infrastructure which includes a possibility of making available data created by the scientific research on demand.

Collecting, organizing and processing research data and facilitating their re-use are activities related to data curation. The term data curation is rather new. Its life began at the beginning of the 1990s when a need for more specific preservation of different types of material or objects had arisen. For instance, museums and libraries need care for physical artefacts while researchers in natural sciences need care for databases comprising data of the human genome.

(Beagrie, 2006, p. 5). For Shreeves and Cragin (2008, p. 89) "Data curation is the active and ongoing management of data through its lifecycle of interest and usefulness to scholarship, science, and education, which includes appraisal and selection, representation and organization of these data for access and use over time". Van Horik (2008, p. 132) explains that there is a subtle difference between data curation and long term digital preservation: "Curation not only implies the preservation and maintenance of a collection or database, but also relates to the creation of added value and knowledge." British Digital curation centre offers similar view of digital curation: "Digital curation involves maintaining, preserving and adding value to digital research data throughout its lifecycle." (Digital curation centre, 2010). According to Lord and Macdonald (2003, p. 12), data curation is: "The activity of, managing and promoting the use of data from its point of creation, to ensure it is fit for contemporary purpose and available for discovery and re-use. For dynamic datasets this may mean continuous enrichment or updating to keep it fit for purpose. Higher levels of curation will also involve maintaining links with annotation and with other published materials.". All these views on data curation suggest existence of an understanding about major issues surrounding data curation, which will lead to finding better solutions for execution of this important activity.

Why is data curation needed?

Data curation is needed for several reasons. Most new researches refer to the findings of the previous researches i.e. data from the previous researches. "Data are evidence supporting research and scholarship; better research is based on verifiable data, which may in turn lead to new knowledge." (Digital curation centre, 2010). By using the output of the previous researches, researchers continue the work of other colleagues who made discoveries before them. That is the main principle how science works. With the growth in volume, complexity, and heterogeneity of digital information, the requirement for active management becomes more challenging and more critical to a wider range of organisations (Van Horik, 2008, p. 133). In addition to these changes, we are witnessing introduction of new electronic devices of different types which provide support to scientists during their research, and produce new research data. As a result, it has become very difficult to collect, organize and process huge amounts of very different types of data and to find purpose for their preservation and re-use. The volume of data created by the scientific research is constantly increasing at a staggering rate, and it has already become huge, so, to secure successful re-use of research data in future, it has become essential to develop technical, organizational, financial and other means to deal with large quantities of research data. In the next decade we will see new experimental facilities coming online that will generate data sets ranging in size from 100's of Terabytes to 10's of Petabytes per year." (Hey and Trefethen, 2003, p. 3). For instance, the Large

Hadron Collider (LHC) at CERN (Geneva) generates roughly 15 petabytes of data annually from 2007 (Van Horik, 2008, p. 133). These volumes of research data are the reason why we need a well planned activity of preservation which would facilitate their future use. Properly curated research data in digital format can be then readily integrated into a new research and learning workflows now and in the future. The research community will benefit from research data curation in four main ways: "1. Improving access. Digital curation procedures allow continuing access to data and improve the speed of access to reliable data and the range of data that can be accessed. 2. Improving data quality. Digital curation procedures assist in improving data quality, improving the trustworthiness of data, and ensuring that data are valid as formal record (such as the use as legal evidence). 3. Encouraging data sharing and reuse. Digital curation procedures encourage and assist data sharing and use by applying common standards and by allowing data to be fully exploited through time (thus maximizing investment) by providing information about the context and provenance of the data. 4. Protecting data. Digital curation procedures preserve data and protect them against loss and obsolence.". (Harvey, 2010, p. 12). Joint Information Systems Committee (JISC) (E-Science Data Curation, 2004) enumerated "(…) seven reasons to keep data: re-use of data for new research, including collection-based research to generate new science; retention of unique observational data which is impossible to re-create; retention of expensively generated data which is cheaper to maintain than to re-generate; enhancing existing data available for research projects; for compliance with legal requirements; to validate published research results and for use in teaching".

Not all data created by scientific research will have long term value. Those data that will have long term value must meet several conditions (Van Horik, 2008, p. 139): "1. Digital research data must be findable by means of a catalogue on the internet. This makes appropriate documentation of the research data relevant. 2. Digital research data must be accessible, provided that privacy rules and intellectual property conditions are taken into consideration. The ultimate goal is to realise open access to the research data, 3. Digital research data must be available in a useable data format, enabling secondary analysis in the future. Therefore the research data must be available in a format that can be processed by common available hardware and software, now and in the future. 4. Digital research data must be reliable, that is, the research data is authentic and not changed in the course of time. 5. Digital research data must be referable in a durable manner. This implies that the research data is provided with persistent identifiers and stored in a in a so-called trusted digital repository.".

During his research related to research data curation, Harvey (2010, p. 56), concluded that "(…) digital curation aims to produce and manage data in ways that ensure they retain three characteristics: longevity, integrity and accessibility. Longevity refers to the availability of the data for as long as their current and future users require them; integrity refers to the authenticity of data – that

411

they have not been manipulated, forged or substituted; authenticity requires that we can locate and use the data in the future in a way that is acceptable to their designated community". These and other authors contributed to the growing body of knowledge about the development of infrastructure for modern networked science and its particular parts which are of great importance to all members of the global scientific community, but also to the society which benefits directly from the scientific endeavour. As science continues to develop, issues like research data curation will attract even more attention than today. The next part of the paper focuses on taking the responsibility for research data curation.

Whose responsibility is it?

Keeping the research data safe is a very responsible job. Scientific community is currently seeking individuals (less likely to take this role), government institutions (more likely to take this role) and commercial enterprises (also more likely to take this role) which are willing to take the role of data keepers.

Scientists are the most responsible for the created research data as they are responsible for carrying out of research projects which produce the research output including data from different measurements, observations and experiments. In the conclusion of her report on dealing with research data, Lyon puts focus on institutional and human aspects of data curation activities. Concerning scientists, the human aspect of data curation, Lyon (2007, pp. 59-60) claims that "(...) many researchers appear to be unaware of the range of issues associated with data management best practice and there is a growing requirement for coordinated advocacy, training and skills programmes to equip the research community with the appropriate competencies to foster the Science Commons envisaged for the future.". Naturally, there are differences among researchers. Jones (2003, p. 3) distinguished three different categories of researchers regarding data curation: "(…) those who don't trust the ability of digital repositories to take care of their material; those who are unaware that such a possibility exists; and those who would love to be able to hand over their materials but no obvious repository yet exists for them to do so.". Generally, to help research data capturing, preservation and exchange, scientists should do the following: apply open-source software and open standards to encourage interoperability among different software and hardware platforms; create metadata and annotations so that digital objects can be reused; link related research materials and make sure the links are persistent; use persistent identifiers; be consistent about citation formats; decide which digital objects need to be curated over longer term, keep data storage devices current; validate and authenticate migrated data. (Harvey, 2010, p. 58). There are many incentives for sharing data with other scientists: publicising the results of their research which occasionally includes data, requirements made by publishers that data underlying an article is to be

made available on request to other researchers, agreements that require data sharing with other research projects in the same area, adherence to open access principles etc. (Ruusalepp, 2008, p. 5).

Because of their position in the research community, academic and research libraries are frequently mentioned as institution where results of the scientific endeavour are kept. Shearer and Argaez (2010, p. 3), suggest that libraries are well positioned to support research data stewardship: "They are already recognized on campus for preserving and providing access to other types of content; and they have strong links with the disciplinary communities.". In case of full text articles (published in scientific journals), books and conference proceedings, libraries nowadays rarely keep local copies of these publications. Instead, they offer access to remote network locations where library users can access and use scientific information resources. As a rule, libraries do not keep research data as part of their holdings. However, the times have changed, and universities would like libraries to become keepers of the local research output. For Angevaare (2009, p. 7) "Libraries have at least three crucial attributes which make them uniquely positioned to curate the output of academic research: 1. they have a mission that includes long-term preservation; 2. they have structural funding; 3. they have a network in the research community.". The question is whether the academic and research libraries are capable of data curation as this type of activity requires constant funding, trained library staff, adequate hardware and software, written data curation policies, logistics support from the academic institution to which library belongs etc. For some libraries, research data curation is an unwanted activity and they reluctantly accept this task. Joint (2007, p. 452) pointed out that „Scientists value their raw data more than the bibliographic expression of that data, and view the preservation of raw data as the prime curational challenge for the knowledge professions such as librarians and archivists.".

If libraries are not going to take care of research data, then we should think about new type of organizations which role would be providing support to e-science by integrating different research material in digital format including datasets of previous research. National Science Foundation (NSF, 2007.) calls for "(...) new type of organizations which will integrate library and archival sciences, cyberinfrastructure, computer and information sciences, and domain science expertise to: provide reliable digital data preservation, access, integration, and analysis capabilities over a decades1long timeline; continuously anticipate and adapt to changes in technologies and in user needs and expectations; engage at the frontiers of computer and information science and cyberinfrastructure with research and development to drive the leading edge forward; and serve as component elements of an interoperable data preservation and access network.". In addition to university libraries, research data are kept by organizations which purpose is to support scientific research and to preserve the output of scientific endeavor. One such example is Digital curation centre

413

(http://www.dcc.ac.uk) in Great Britain. This organization is "(…) gateway to the technical solutions, curation tools and learning resources that can help data custodians like you to build capacity for digital curation." (Digital curation centre, 2010). These and other similar organizations (among which some are working for profit) will have a significant role in determining the future development of science in general. As McGovern and McKay (2008, p. 262) pointed out, the biggest problem will be finding organizations that "(…) can sustain their commitment to archive the digital content (…)".

These and other suggestions about who will take the responsibility for long term preservation of output of scientific research will help in discovering pros and cons of each possible solution offered. No final decision has been made yet about what organization will take the responsibility for data curation. At the moment, this responsibility is shared between scientists themselves, libraries and different organizations which curate data on commercial bases for profit. The principal question is who among them will demonstrate more dedication for preservation of research output in the long run?

Where research data are to be kept?

Digital curation needs reliable digital information systems that guarantee long term preservation, integrity and accessibility of research data. Digital repositories are most recent type of information resources that emerged in the academic community in the first half of the 1990s. Digital institutional repository (a digital information repository that is a part of university or other institution) is a digital archive of the intellectual product created by the faculty, research staff, and students of an institution and accessible to end users both within and outside of the institution, with few if any barriers to access (Johnson, 2002). An institutional digital repository can contain e-prints of scientific papers, research data, but also e-learning materials and other forms of institutional intellectual outputs, which are generally not published or preserved elsewhere (Hockx-Yu, 2006, pp. 234-235). Digital institutional repositories have a great importance in preservation of the scientific output: they focus organizational attention on managing digital content; they provide a potential entry point even a back door—for getting content into digital preservation programs; depositors and other stakeholders in institutional repositories may learn about digital preservation issues when they deposit digital content into institutional repositories; institutional repositories may offer an opportunity to address preservation planning priorities by providing guidelines and tools for depositors to prepare archive-ready digital content and they preserve retiring faculty's digital legacy (McGovern and McKay, 2008, p. 268). Digital repositories can store different file formats and types of content. Since data formats can become obsolete, repositories should only accept data in standardized formats. So far digital repositories have proved to be the best possible solution for the problem of data cu-

ration in this transitional period. Digital repositories are also the core of many projects related to data curation. Some examples of such projects are: Digital Curation Centre at http://www.dcc.ac.uk, DARIAH at http://www.dariah.eu/, University of Minnesota project at http://www.lib.umn.edu/datamanagement/ archiving, National Geological and Geophysical Data Preservation Program at http://datapreservation.usgs.gov/index.shtml etc. There are also organisations like Datacite at http://datacite.org/ which help establish easier access to research data. Both projects and organisations help make understanding of data curation process better and help scholars and all other interested parties to get involved in this important activity.

Conclusion

"Maintenance of a complete and accurate scholarly record, including the portion in digital form, is essential for continued progress in research and learning." (Beagrie, Chruszcz, and Lavoie, 2008, p. 16). We can agree that this idea is one of the most important ideas that will support the development of science in the future by taking care of the output of its previous research so it can be integrated into new scientific research. As Isaac Newton wrote in 1676 "If I have seen a little further it is by standing on the shoulders of Giants." (The Phrase Finder, 2010). Today, a significant attention is given to the preservation of printed (as well as digital) scientific output such as journal articles, books, and conference papers published in proceeding, and less attention is given to the preservation of research data. Research data has important role in science as it can provide support to new scientific research. To preserve research data, a new type of activity called data curation has been developed. The purpose of this activity is to facilitate re-use of data of previously completed scientific research in a new scientific research. According to Van Horik (2008, p. 132) "Digital curation or data curation is needed to maintain digital materials, such as research data, over their entire life cycle and over time for current and future generations of users.". For this activity to be successful, science needs adequate infrastructure which will enable collecting, storing, organizing and re-use of research data. With development of that infrastructure, the science will be able to access more easily its fundaments and make new scientific discoveries possible.

References

Angevaare, Inge. Taking Care of Digital Collections and Data: 'Curation' and Organisational Choices for Research Libraries. Liber Quarterly. 19(2009), 1; 1-12.

Digital Curation Centre. http://www.dcc.ac.uk (3.6.2011.)

E-Science Data Curation. http://www.jisc.ac.uk/publications/generalpublications/2004/pub_escience.aspx. (1.5.2011.)

Harvey, Ross. Digital Curation : a how-to-do-it manual. New York, London : Neal-Schuman Publishers, 2010.

Hey, Tony; Trefethen, Anne. The Data Deluge: An e-Science Perspective. // Grid Computing—Making the Global Infrastructure a Reality. New York: John Wiley, 2003 http://eprints.ecs.soton.ac.uk/7648/1/The_Data_Deluge.pdf (5.4.2011.)

Hockx-Yu, Helen. Digital preservation in the context of institutional repositories. // Program: electronic library and information systems. 40(2006), 3; 234-245.

Ruusalepp, Raivo. Infrastructure Planning and Data Curation: A Comparative Study of International Approaches to Enabling the Sharing of Research Data. JISC, 2008. http://www.dcc.ac.uk/sites/default/files/documents/publications/reports/Data-Sharing-Report.pdf (10.6.2011.)

Joint, Nicholas. Data preservation, the new science and the practitioner librarian. // Library Review. 56(2007), 6; 450-455.

Jones, Maggie. Digital Preservation Activities in the U.K – building the infrastructure. http://archive.ifla.org/IV/ifla69/papers/129e-Jones.pdf (9.2.2011.)

Johnson, Richard K. Institutional repositories: partnering with faculty to enhance scholarly communication. // D-Lib Magazine. 8(2002), 11, http://www.dlib.org/dlib/november02/ johnson/ (1.5.2011.)

Lyon, Liz. Dealing with data: roles, rights, responsibilities and relationships. http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf (9.2.2011.)

Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century. Arlington: National Science Foundation, 2005.

Lord, Philip.; Macdonald, Alison. Data curation for e-science in the UK: an audit to establish requirements for future curation and provision: e-Science Curation Report. http://www.jisc.ac.uk/uploaded_documents/e-ScienceReportFinal.pdf (9.2.2010.).

McGovern, Nancy Y.; McKay, Aprille C. Leveraging Short-term Opportunities to Address Long-term Obligations: A Perspective on Institutional Repositories and Digital Preservation Programs. // Library Trends. 57(2008), 2; 262-279.

Ogburn, Joyce L. The Imperative for Data Curation. // Libraries and the Academy. 10(2010), 2; 41–246.

Shearer, Kathleen; Argáez, Diego. Addressing the Research Data Gap: A Review of Novel Services for Libraries. Ottawa: CARL, 2010.

Shreeves, Sarah L.; Cragin, Melissa H. Introduction: Institutional Repositories: Current State and Future. // Library Trends. 57(2008),2 ; 89-97.

Sustainable Digital Data Preservation and Access Network Partners (DataNet), National Science Foundation, 2007 http://www.nsf.gov/pubs/2007/nsf07601/nsf07601.htm (28.4.2011.)

Van Horik, Rene. Data curation. // A Driver's Guide to European Repositories. Amsterdam : Amsterdam University press, 2008. 131-152.

The Phrase Finder http://www.phrases.org.uk/meanings/268025.html (10.6.2011.)

Best Archival Practice in the Regulation of Medicines: Work on the Guidelines for Agencies for Medicinal Products

Arian Rajh
Agency for Medicinal Products and Medical Devices
Ksaverska cesta 4, Zagreb, Croatia
arian.rajh@halmed.hr

Hrvoje Stančić
Department of Information Sciences,
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
hrvoje.stancic@zg.t-com.hr

Summary

This paper announces the beginning of the work on the standardization of filing of electronic medicinal dossiers in agencies for medicinal products. Medicinal products electronic dossiers have to be preserved for the long-term because of the requirements for protection of patients' safety and standardization of archiving procedures.

Key words: common European best archival practice, electronic Common Technical Document, eCTD, Open Archival Information System

Introduction

Management and long-term preservation issues connected with the medicinal products' documentation became a very important topic since the end of 2009 – the year when the European agencies for medicinal products were supposed to switch from the paper-based records to the electronic records. Since the field of medicines' regulation and control is constantly being challenged for pharmacovigillance and legal reasons, it is of the utmost importance to standardise the procedures of management, storage and long-term preservation of electronic records in this domain.

Having all said in mind, a new subgroup with the mandate of creation of guidance for prescribing archival procedures for medicinal products' electronic submissions in the EU national agencies for medicinal products and the European medicines agency (EMA) is found by the Telematic Implementation Group on electronic submission (TIGes). Electronic submission denotes electronic medicinal products' documents, i.e. basic documentary resource for me-

dicinal products regulation and control. Although relevance of such guidance was emphasised before the foundation of the subgroup, now more and more European agencies for medicinal products are joining this initiative for preparing the guidance and proposing it to the TIGes.

TIGes is one of the telematic projects[1] that gather European agencies for medicinal products. TIGes' meetings are hosted by EMA in London. Telematic projects are way to implement common European strategy of working with electronic resources for regulation of medicines since 2003, and TIGes group is responsible for implementation of electronic submissions since 2000. Beside electronic submissions, telematic projects involve managing of common European database for clinical trials, manufacturing and manufacturers data, database of medicinal products for human on the European market, common human pharmacovigillance system, veterinary pharmacovigillance database, and managing infrastructure and referral projects, such are the system for safe exchange of data between agencies for medicinal products (EudraLink), secure European regulatory network of EMA and national agencies for medicinal products (EudraNet), European review system for assessments of electronic submissions (EURS), and telematics controlled terms repository (EUTCT).

TIGes group is in charge for production, approval, implementation and monitoring of all European standards and guidelines that support European and global specifications for electronic submission in the electronic common technical document format (eCTD) and guidelines for additional European non-eCTD electronic submission format. TIGes meets four times a year, the first day is reserved for meeting of regulatory delegates just from agencies for medicinal products, and second day is for meeting of regulatory delegates with pharmaceutical industry associations representatives and other invited guest, according to current topics and business needs that need to be discussed and solved. There was forty six meetings held until now and during this eleven years period various subgroups were established, such are subgroup for implementation of veterinary electronic submissions (TIGes Vet), subgroup for interlinking of various standards and their implementation (Interlinking), subgroup for implementation of European review system specification published by EMA and monitoring various vendors' tools (EURS), subgroup for development of central repository of electronic submissions received by EMA in centralised procedure, subgroup for development of common European portal for receiving electronic submissions for all agencies, subgroup for production of electronic application (submission) form (eAF), and subgroup for (medicinal) product information management (PIM). It can be stated that TIGes represents a contemporary dynamic work environment adaptable to implementation of telematic strategy and management of different projects, e.g. while further work of PIM and EURS sub-

---

[1] Introduction to the EU telematics programme, 19 May 2011, http://www.ema.europa.eu/docs/en_GB/document_library/Other/2011/05/WC500106712.pdf, June 2011.

groups remains uncertain because of closing of PIM project and change of EURS mandate, new subgroups like *ad-hoc* guidance group Best archiving practice guidance is being born.

Best archiving practice and guidance

Ad-hoc guidance group Best archiving practice is a subgroup of TIGes with the mandate of creation of common European best archiving practice document. The goal of long-term preservation attempts in this case is medicinal products' dossiers, so called "electronic submissions". An electronic submission consists of records on manufacturing a medicinal product and its characteristics, substances and ingredients. These records are compiled into parts I-IV according to an older format or into modules 1-5 according to the current Common Technical Document (CTD) format. *Ad-hoc* guidance group Best archiving practice was founded upon proposition of Cécile Lombard from French agency for medicinal products and Arian Rajh from Croatian counterpart-agency after the visit of Croatian delegation to the French colleagues in January 2011. Additional reason to found this particular subgroup lied in the results of an internal TIGes' survey on the EU agencies' archives. The 2010 TIGes' survey on archives showed that the EU agencies for medicinal products did not established coherent archival policies, processes and systems capable of long-term preservation of electronic submissions (or electronic medicinal products dossiers). The provisional idea of solving the problem of long-term preservation was presented to TIGes' delegates as a business case which has been translated into the mandate of the proposed *ad-hoc* guidance group in the later stage. *Ad-hoc* guidance group is chaired by the Cécile Lombard and current members are delegates from Poland, Denmark, France, Croatia, Hungary, Finland, Luxembourg, Malta, Austria, the Nederland, Greece, Slovenia and Estonia. It is expected that the delegates from Sweden and EMA join the *ad-hoc* guidance group. EMA ensures support and logistic for web-conferences of this *ad-hoc* group.

Guidance should include mutually recognised requirements for all agencies and their information systems, and recommendations for establishment of national and organisation-specific solutions for long-term archiving of electronic submissions. These mutually recognised requirements and organisation or nationally-oriented recommendations should enhance long-term preservation framework. The future guidance will refer to the following existing specifications, guidelines and standards:

- European specification for administrative module of electronic eSubmissions (current version is 1.4. from 2009)
- Modules 2-5 of electronic submissions global specification (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, ICH)

- Electronic submission guidelines for eCTD i NeeS format issued by TIGes
- General International Standard Archival Description standard ISAD(G) and International standard archival authority record for corporate bodies, persons and families ISAAR(CPF)
- ISO 14721:2003 Space data and information transfer systems – Open archival information system – Reference model
- ISO 15489-1:2001 Information and documentation – Records management – Part 1 and Part 2
- ISO 23081 Information and documentation – Records management processes – Metadata for records – Part 1 and Part 2
- Model Requirements for Electronic Records Management specification

OAIS and the referent framework of the guidance

The referent framework for archiving electronic submission in eCTD format and records submitted to the agencies in support of regulation of medicines should include mutual requirements, stated by all agencies, and more-loose recommendations that could fit in the broader national and organisational practices. An information model provided in ISO standard for Open Archival Information System (OAIS) should be used as the foundation of such highly productive referent framework.

Open Archival Information System is a standardised reference model of an archive, made of human organisation and systems, capable of accepting responsibility for preserving information and for making it available for designated community. In this case openness means possibility of various implementations of the information model, not unlimited access.[2] OAIS is also the reference model for information systems capable for long-term preservation of digital contents. OAIS prescribes information and functional model. Fundamental entity of OAIS information model is information object. Information is defined in OAIS standard as a combination of set of bits and their representation. Information object is compiled as data object and representation information which makes possible to interpret set of bits of the data object. We are referring to the digital data objects, although the reference model encompasses physical objects, too. Representation information ensures interpretation of digital data object by providing its meaning. There are several types of representation information: structure information, semantic information or representational network. Information objects are divided to content information, created when representation information interprets content data object, preservation description information, packaging information or descriptive information. Basic information objects and primary goal of preservation is the content information. In order to preserve

---

[2] ISO 14721:2003, 1-11.

it the system generates archival information package. Packaging information aims to bond content information, preservation description information and descriptive information. These information are rapt together into information package which can be archival (AIP), dissemination (DIP) or submission information package (SIP). Submission information package is input into OAIS system and disseminated package is part of the content provided for users. ISO OAIS standard also defines basic functions of the OAIS system/archive. These functions should be implemented by Best archival practice for electronic submissions and they should deal with execution of complex verifications of packages due to the risk of content obsolescence and preparation for migrations.

Problems related to electronic submissions

Each part or module contains particular medicinal product information. It can be administrative information, (pharmeceutical-chemical) quality related information, nonclinical and clinical information. Quality information is used for assessment and testing of quality of medicines, nonclinical information for efficacy and clinical information for approbating of safety of medicines. Nonclinical documents contain information provided by the medicinal product's manufacturer and scientists conducting approbation of pharmacological profile of the product, and pharmacocinetics, pharmacodynamics and toxicology tests. Common technical document (CTD) paper format, dividing dossiers into modules, have one additional module – module 2 or summary of the dossier. CTD format has also electronic version – electronic Common technical documents format (eCTD). eCTD dossier contains PDF files within structured folders. For lifecycle of the dossier there is XML DTD in "util" folder of the dossier. Each dossier is published by the medicinal product manufacturer and submitted to the medicinal product agency as a sequence, so dossier can be seen as a set of sequences counting from 0000 onwards. Granulation of folders and filenames in sequences are strictly specified. Metadata is extracted from XML and used by the review systems or applications for representation of eCTD dossiers. That is how assessment of medicinal products is facilitated because assessors can retrieve the data effectively. New, appended, deleted or replaced documents are emphasised in the review tool with different colour. In Reykjavik in February 2005 Heads of the European medicines agencies set out the end of 2009 as the target date for shifting from the paper based format to eCTD. Today most of the medicinal products agencies work with eCTD but they still have not established filling policies and procedures. Proactive approach for long-term preservation of eCTD dossiers on the European level is something that should be established. The basic problem in working with medicinal product dossiers is complexity of relations between dossier, case and regulatory procedure. That problem exists regardless of format of the dossier, either being paper or electronic. But if an organisation wants effective control of eCTDs and electronic documents in general, it is obligatory to carefully design the organisation's business processes

that use electronic documentation as resource. Misplaced paper files can be found eventually, but maintenance and management of complex digital objects is much more demanding.

What are relations between dossier, cases and regulatory procedures? One dossier can be business resource for many medicinal products (cases), i.e. for capsules and suspension, which can be divided into two separate medicine products sharing same active substance. One dossier can be resource for control of many medicinal products, i.e. when submitted as a dossier for group variation. Regulatory procedures that use dossiers are procedures of registration, renewal of registration and variation of approval. There are several types of regulatory procedures, because medicinal products could be regulated either by the European medicines agency in a centralised procedure or by national agencies for multiple markets (mutual recognition procedure or decentralised procedure) or for their own national markets (national procedure). The same dossier can be resource in more than one procedure, i.e. for variation of solution and for registration of capsules which are new on some national market. Or, dossier once used for registration of medicinal product on common EU market could be used for registration of the same product outside EU. In conclusion, national agencies' for medicinal products function is national administrations and they are obligated to follow registration offices' procedures which distinguish medicinal products, cases and regulation procedures.

The fundamental question in the preparatory work on Best archiving guidance is – what to preserve? In addition to preservation of dossiers, submitted by manufacturers of medicinal products, there is a need to preserve additionally submitted medicinal products-related documents, that do not fit eCTD specification and therefore they are placed outside of the eCTD record, then records created by agencies for medicinal products during assessments, as well as the final version of documents for medicinal personnel and patients.

Conclusion

Creation of the Best archiving practice guidance will be very demanding due to the complexity of structure and relation of digital objects that we want to preserve over long term. One of the most useful experiences that can facilitate that attempt is a prototype application for packaging eCTD dossiers and other records into OAIS-compliant archival information packages, presented to the public on eTELEMED 2011[3] international conference. The prototype showed possibilities of extracting metadata from the content, adding preservation metadata, mapping different metadata schemes, packaging content, protection of packages, and analysis of potentially obsolete parts of the content.

---

[3] Stančić, Hrvoje, Rajh, Arian, Pavlina, Krešimir. Long-term Preservation Solution for Complex Digital Objects Preserved as Archival Information Packages. eTELEMED 2011, 23-28.02. 2011., Gosier, Guadeloupe, France.

Once the creation process of the Best archiving practice starts, it will be important to define critical components of archival information packages and logical connections between them. Furthermore, it can be assumed that versioning of packages could be an issue and protection of every version could be yet another requirement. It should be defined which metadata to preserve and how to structure them. It will also be necessary to test packages periodically in order to detect possible obsolescence of content or file formats contained in packages, to analyse components, to prepare packages for migrations, to version packages with the preservation-related changes and to protect that new versions. The final goal should be to have the possibility to retrieve and use archival information packages for particular medicinal product or products that have been sold on particular market in a particular time period.

Beside the expected usage of the Best archiving practice by agencies for medicinal products it could also be used by pharmaceutical companies. The pharmaceutical companies have difficulties to answer questions like: "Which dossier was submitted to which agency?", "Which documents were valid in which time?" etc. because of the rising amounts of dossiers, number of agencies that regulate their medicines, and complexity of regulation procedures. These questions are usually asked for pharmacovigilance and legal reasons, and agencies must be able to answer them quickly and correctly as well as be able to prove the authenticity, integrity and usability of the preserved electronic records at the same time making no room for questioning their trustworthiness.

References

Gladney, H. M. LongTerm Preservation of Digital Records, Part I: A Theoretical Basis, ERPAePRINTS, October 2008, http://eprints.erpanet. org/152/ June 2011.

ICH M2 EWG electronic common technical document specification, July 2008, http://estri.ich. org/ectd/eCTD_Specification_v3_2_2.pdf June 2011.

Introduction to the EU telematics programme, 19 May 2011, http://www.ema.europa.eu/ docs/en_GB/document_library/Other/2011/05/WC500106712.pdf, June 2011.

ISAD(G): General International Standard Archival Description standard, International Council On Archives, Second edition, 2000, http://www.icacds.org.uk/eng/ISAD%28G%29.pdf, June 2011.

ISAAR (CPF): International Standard Archival Authority Record For Corporate Bodies, Persons and Families, International Council On Archives, Second edition, 2003, http://www.icacds. org.uk/eng/ISAAR%28CPF%292ed.pdf

ISO 14721:2003 Space data and information transfer systems – Open archival information system – Reference model.

ISO 15489-1:2001 Information and documentation – Records management – Part 1 and Part 2.

ISO 23081 Information and documentation – Records management processes – Metadata for records – Part 1 and Part 2.

Model Requirements for Electronic Records Management specification (MoReq2010), http://moreq2010.eu/pdf/MoReq2010-Core+Plugin(v1-0).pdf, June 2011.

Stančić, Hrvoje, Rajh, Arian, Pavlina, Krešimir. Long-term Preservation Solution for Complex Digital Objects Preserved as Archival Information Packages. eTELEMED 2011, 23-28 February 2011, Gosier, Guadeloupe, France.

INFuture2011 and KEEP Project
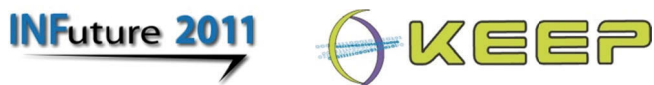
INFuture2011 conference has joined forces with KEEP Project in order to promote the importance of long-term preservation issues concerning digital materials. During the conference, KEEP Project's partners will disseminate their project results through the series of self-contained sessions, all together in the form of workshop.

About KEEP Project
Coordinating organisation of the project is:
- Bibliothèque nationale de France

Other project partners are:
- Joguin SAS, France
- Koninklijke Bibliotheek, The Netherlands
- Computerspiele Museum, Germany
- University of Portsmouth, United Kingdom
- Deutsche Nationalbibliothek, Germany
- Cross Czech a.s., Czech Republic
- Tessella, United Kingdom
- European Games Developer Federation, Germany

KEEP (Keeping Emulation Environments Portable) aims to develop emulation services (KEEP Emulation Services) to enable accurate rendering of both static and dynamic digital objects: text, sound, and image files; multimedia documents, websites, databases, videogames etc.

The overall aim of the project is to facilitate universal access to our cultural heritage by developing flexible tools for accessing and storing a wide range of digital objects. KEEP will also consider legal issues concerning the implementation of emulation-based systems and propose solutions which comply with European and national copyright laws.

The very success of computing technology, where machines are rapidly superseded, has created a serious and growing challenge of how to preserve access to digital material produced on obsolete machines. Cultural heritage organisations are particularly sensitive to the threat of major data loss resulting from technical obsolescence. KEEP will develop the KEEP Emulation Services to enable the accurate rendering of these objects, designed for a wide variety of computer systems, so that they can be securely accessed in the long term.

KEEP will address the problems of transferring digital objects stored on outdated computer media such as floppy discs onto current storage devices. This will involve the specification of file formats and the production of transfer tools exploited within a framework, and taking into account possible legal and technical issues. KEEP will address all aspects ranging from safeguarding the original bits from the carrier to offering online services to end-users via a highly portable emulation framework running on any possible device. In addition to producing a software package, the project will deliver understanding about how to integrate emulation-based solutions with an operational electronic deposit system. Existing metadata models will be researched and guidelines will be developed for mapping digital objects to emulated manifestations. KEEP will seek ways to integrate its work with the outputs of other digital preservation projects and software (for example Planets and Pronom). Overall, KEEP will contribute to the next generation of permanent access strategies based on emulation.

Although primarily aimed at those involved in Cultural Heritage, such as memory institutions and games museums, the KEEP Emulation Services can also serve the needs of a wide range of organisations and individuals because of its universal approach.[1]

KEEP sessions during the INFuture2011 conference

1. Introduction to the Tools and Services of the KEEP Project, including Business Case Drivers and Legal Issues

Members of the KEEP Project will provide an introduction to the project, explaining its purposes and how its tools and services provide a compelling Business Case for the use of emulation for certain aspects of Digital Preservation.

There will also be an introduction to some of the legal issues which arise under international and European Law which have an effect on Digital Preservation.

2. Explanation of and hands-on workshop with the KEEP Emulation Framework

A great many computer emulation programmes exist which might assist with access to old software and data on a variety of platforms. But these can sometimes require considerable expertise to select and configure before they can be used.

The KEEP Emulation Framework has automated much of these activities into a set of open tools which are much more user-friendly.

During this session, the KEEP Team will demonstrate, and then invite delegates to experiment with the tools in the Framework.

---

[1] About KEEP, http://www.keep-project.eu/ezpub2/index.php?/eng/About-KEEP

3. Explanation of and hands-on workshop with the KEEP Data Transfer Framework and Mediabase

Before old data can be accessed, it might be necessary to transfer it from old and obsolete data carriers – for example 5.25-inch floppy disks.

In this session, the KEEP Team will explain how their Data Transfer Framework will operate – both stand-alone and also as a service with the PLANETS suite of tools and services.

There will also be a demonstration of the KEEP Mediabase, which delegates will be able to experiment with themselves.

4. Metadata Considerations in Emulation and TOTEM database

Although issues of Metadata in Digital Preservation are well-known and much-written about, different requirements arise when planning to retain the ability to emulate old systems. In this session, the KEEP Team will consider this topic in more detail and will introduce delegates to the TOTEM database, designed within KEEP to identify the components necessary to access different types of data and also the relationships between different hardware, software and operating systems. Delegates will have the opportunity to experience this tool through hands-on training and will retain access to it, via the Internet, after the end of the conference.

5. Planning for the future use of emulation using the KEEP Virtual Machine

The KEEP Virtual Machine is an innovative piece of research which has produced software enabling emulation to provide a sustainable means to bridge the changes in hardware and operating systems which occur all too frequently. The KEEP Team will provide a detailed explanation of the technical approach which has been taken, and will demonstrate current prototypes. Delegates will have an opportunity to take part in some hands-on experimentation.

INFuture Conference Standing Board of Reviewers

Vjekoslav Afrić, Faculty of Humanities and Social Sciences, University of Zagreb

John Akeroyd, South Bank University, London

Yvette M. Alex-Assensoh, Indiana University Bloomington

Silvija Babić, Croatian State Archives, Zagreb

Miroslav Bača, Faculty of Organization and Informatics, University of Zagreb Varaždin

Mihaela Banek Zorica, Faculty of Humanities and Social Sciences, University of Zagreb

Ana Barbarić, Faculty of Humanities and Social Sciences, University of Zagreb

Anabela Barreiro, Institute for Systems and Computer Engineering: Research and Development – INESC-ID, Lisboa

Bob Bater, InfoPlex Associates Ltd., Bristol

David Bawden, City University, London

Zoran Bekić, University Computing Center, Zagreb

Inoslav Bešker, University of Bologna

Clive Billenness, The British Library

Helen Boelens, The Netherlands

Goran Bubaš, Faculty of Organization and Informatics, University of Zagreb Varaždin

Viktorija Car, Faculty of Political Science, University of Zagreb

Andrew Cox, University of Sheffield

Blaženka Divjak, Faculty of Organization and Informatics, University of Zagreb

Senada Dizdar, Faculty of Philosophy, Sarajevo

Milena Dobreva, University of Portsmouth

Zdravko Dovedan Han, Faculty of Humanities and Social Sciences, University of Zagreb

Zvjezdana Dukić, Hong Kong

Annette Đurović, Faculty of Philology, University of Belgrade

Karmen Erjavec, Faculty of Social Sciences, University of Ljubljana

Aleš Gačnik, Znanstveno-raziskovalno središče Bistra, Ptuj

Franz Guenthner, Centrum für Informations- und Sprachverarbaitung, Ludwig-Maximilians-Universität München

Ivana Hebrang Grgić, Faculty of Humanities and Social Sciences, University of Zagreb

Anette Hemera, CSoft International, Budimpešta

Aleksandra Horvat, Faculty of Humanities and Social Sciences, University of Zagreb

Maja Jokić, Institute for Social Research, Zagreb

Jelena Jurišić, Croatian Studies, University of Zagreb

Lejla Kodrić Zaimović, Faculty of Philosophy, Sarajevo
Steven Krauwer, Utrecht institute of Linguistics UiL-OTS, The Netherlands
Vlasta Kučiš, University of Maribor
Stefan Langer, Centrum für Informations- und Sprachverarbaitung, Ludwig-Maximilians-Universität München
Jadranka Lasić-Lazić, Faculty of Humanities and Social Sciences, University of Zagreb
Nikolaj Lazić, Faculty of Humanities and Social Sciences, University of Zagreb
Marko Lukičić, Ericsson Nikola Tesla, Zagreb
Maurizio Lunghi, Fondazione Rinascimento Digitale, Florence
Stjepan Malović, University of Dubrovnik
Luisa Marquardt, Roma Tre University
Vladimir Mateljan, Faculty of Humanities and Social Sciences, University of Zagreb
Nives Mikelić Preradović, Faculty of Humanities and Social Sciences, University of Zagreb
Anja Nikolić Hoyt, Croatian Academy of Sciences and Arts
Đorđe Obradović, University of Dubrovnik
Aleksander Panjek, University of Primorska, Koper
Krešimir Pavlina, Faculty of Humanities and Social Sciences, University of Zagreb
Jelka Petrak, School of Medicine, University of University of Zagreb
Nenad Prelog, University of Zadar
Arian Rajh, Agency for Medical Products and Medical Devices Zagreb
Andreas Rauber, Vienna University of Technology (TU-Wien)
Sanja Seljan, Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Simeon, Faculty of Humanities and Social Sciences, University of Zagreb
Karolj Skala, Ruđer Bošković Institute, University of Zagreb
Aida Slavić, UDC Consortium
Sonja Špiranec, Faculty of Humanities and Social Sciences, University of Zagreb
Hrvoje Stančić, Faculty of Humanities and Social Sciences, University of Zagreb
Marko Tadić, Faculty of Humanities and Social Sciences, University of Zagreb
Miroslav Tuđman, Faculty of Humanities and Social Sciences, University of Zagreb
Ludmila Ticha, Czech Technical University in Prague
Kostandina Veljanovska, Faculty of Administration and Managment of Information Systems, St. Kliment Ohridski University, Bitola
Radovan Vrana, Faculty of Humanities and Social Sciences, University of Zagreb
Neven Vrček, Faculty of Organization and Informatics, University of Zagreb Varaždin

Aleksandra Vraneš, Faculty of Philology, Belgrade University
Kristina Vučković, Faculty of Humanities and Social Sciences, University of Zagreb
Žarka Vujić, Faculty of Humanities and Social Sciences, University of Zagreb
Mirna Willer, Department of Library and Information Science, University of Zadar
Goran Zlodi, Faculty of Humanities and Social Sciences, University of Zagreb
Daniela Živković, Faculty of Humanities and Social Sciences, University of Zagreb

INFuture2011 conference was sponsored by

Ericsson Nikola Tesla d.d.
http://www.ericsson.com/hr/

SDL Zagreb d.o.o.
http://www.sdl.com

INFuture2011 conference cooperated with

KEEP (Keeping Emulation Environments Portable) Project
http://www.keep-project.eu